VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA OSTRAVA
HORNICKO-GEOLOGICKÁ FAKULTA
INSTITUT GEOINFORMATIKY

Prostorové shlukování v městském prostředí

DIPLOMOVÁ PRÁCE

Autor práce: Bc. Pavel Soukup
Vedoucí práce: doc. Dr. Ing. Jiří Horák

Ostrava 2011

VŠB – TECHNICAL UNIVERSITY OF OSTRAVA
FACULTY OF MINING AND GEOLOGY
INSTITUTE OF GEOINFORMATICS

Spatial clustering in urban areas

DIPLOMA THESIS

Author: Bc. Pavel Soukup
Supervisor: doc. Dr. Ing Jiří Horák

2011

VŠB - Technická univerzita Ostrava
Hornicko-geologická fakulta
Institut geoinformatiky

Zadání diplomové práce

Student: **Bc. Pavel Soukup**

Studijní program: N3646 Geodézie a kartografie

Studijní obor: 3602T002 Geoinformatika

Téma: Prostorové shlukování v městském prostředí
Spatial Clustering in Urban Areas

Zásady pro vypracování:

Úkoly:
1) Aspekty prostorové segregace a její hodnocení pomocí prostorových indikátorů
2) Rešerše metod shlukování
3) Návrh postupu pro realizaci prostorového shlukování
4) Zpracování, testování a vyhodnocení na vybraných případech sociálních-ekonomických a demografických dat pro Ostravu
5) zpracování resumé ve světovém jazyce o rozsahu nejméně 3 strany A4.

Seznam doporučené odborné literatury:

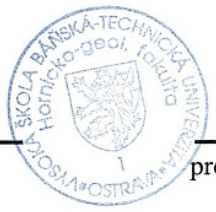
Horák J.: Prostorová analýza dat. Skripta VŠBTUO, 2006. 149 stran.
JARGOWSKY, Paul, A. Concentration of Poverty and Metropolitan Development. 2006. University of Texas at Dallas.
MELOUN, Milan; MILITKÝ, Jiří; HILL, Martin. Počítačová analýza vícerozměrných dat v příkladech. Vydání 1. Praha : Nakladatelství Akademie věd České republiky, 2005. 449 s. ISBN 80-200-1335-0.
STILLWELL, J., CLARKE, G.: Applied GIS and Spatial Analysis. John Wiley & Sons, Ltd, 2006.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **doc. Dr.Ing. Jiří Horák**

Datum zadání: 31.10.2010

Datum odevzdání: 30.04.2011



prof. Ing. Zdeněk Diviš, CSc.
vedoucí institutu

prof. Ing. Vladimír Slivka, CSc., dr.h.c.
děkan fakulty

PROHLÁŠENÍ

- Celou diplomovou jsem vypracoval samostatně a uvedl jsem všechny použité podklady a literaturu.

- Byl jsem seznámen s tím, že na moji diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. - autorský zákon, zejména § 35 – využití díla v rámci občanských a náboženských obřadů, v rámci školních představení a využití díla školního a § 60 – školní dílo.

- Beru na vědomí, že Vysoká škola báňská – Technická univerzita Ostrava (dále jen VŠB-TUO) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užít (§ 35 odst. 3).

- Souhlasím s tím, že jeden výtisk diplomové práce bude uložen v Ústřední knihovně VŠB-TUO k prezenčnímu nahlédnutí a jeden výtisk bude uložen u vedoucího diplomové práce. Souhlasím s tím, že údaje o diplomové práci, obsažené v Záznamu o závěrečné práci, umístěném v příloze mé diplomové práce, budou zveřejněny v informačním systému VŠB-TUO.

- Souhlasím s tím, že diplomová práce je licencována pod Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licencí. Pro zobrazení kopie této licence, je možno navštívit http://creativecommons.org/licenses/by-nc-sa/3.0/

- Bylo sjednáno, že s VŠB-TUO, v případě zájmu z její strany, uzavřu licenční smlouvu s oprávněním užít dílo v rozsahu § 12 odst. 4 autorského zákona.

- Bylo sjednáno, že užít své dílo – diplomovou práci nebo poskytnout licenci k jejímu využití mohu jen se souhlasem VŠB-TUO, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly VŠB-TUO na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Ostravě dne 28. 4. 2011

.....

Pavel Soukup

ANOTACE

V první části této diplomové práce je popsána prostorová segregace v obecné rovině. V práci jsou zmíněny zejména její formy a aspekty, ale i nástin toho, jak by jí šlo předcházet. V další části jsou rozebrány indexy rezidenční separace a prostorové izolace skupin obyvatel neboli míry segregace. V následujícím úseku práce se zabývám tématem prostorové autokorelace a jejím měřením pomocí Moranova I kritéria a analýzy LISA na území města Ostravy. V neposlední řadě je značná část práce věnována shlukovým analýzám obecně a samozřejmě i jejich aplikacím v urbánním prostředí. Těžištěm práce je porovnání různých metod shlukování.

Klíčová slova:

analýza shluků, prostorová segregace, indexy prostorové heterogenity, prostorová autokorelace, městské prostředí

ANOTATION OF THESIS

In the first part of this thesis there is described the spatial segregation in general. Its forms and aspects, as well as an outline how to prevent it are mentioned in the work. In the second part there are portrayed the residential separation indexes and spatial isolations of population groups or the degree of separation. In the next section I deal with spatial autocorrelation and its measuring with Moran's I criterion and with LISA analysis in the area of the city of Ostrava. Last but not least, a big part of the work is dedicated to the cluster analyses in general and, of course, to the applications of them in the urban areas. The focus of the diploma thesis is the comparison of the different clustering methods.

Keywords:

cluster analysis, spatial segregation, spatial heterogenity indexes, spatial autocorrelation, urban areas

PODĚKOVÁNÍ

Touto cestou bych rád poděkoval svému vedoucímu diplomové práce doc. Dr. Ing. Jiřímu Horákovi za konzultace, odborné vedení, přínosné připomínky a praktické rady při postupu řešení této práce.

OBSAH

SEZNAM ZKRATEK

České zkratky:

ČSÚ – Český statistický úřad

DKM – Digitální katastrální mapa

ISEO – Informační systém evidence obyvatel

ISKN – Informační systém katastru nemovitostí

MMO – Magistrát města Ostravy

OKD – Ostravsko–karvinské doly

PI – Přirozená identifikace

RSO – Registr sčítacích obvodů a budov

S-JTSK – Systém jednotné trigonometrické sítě katastrální

SLDB – Sčítání lidu, domů a bytů

ÚP – Úřad práce

Zahraniční zkratky:

ASCII – American Standard Code for Information Interchange

CCC – Cubic clustering criterion

CLU – Cluster analysis

GAC – Gabal Analysis & Consulting

HH – High value surrounded by high values

LAU – Local Administrative Units

LISA – Local ndicators of spatial association

LL – Low value surrounded by low values

NUTS – Nomenclature des Unites Territoriales Statistique

TCSS – Total cluster sum of squares

VBA – Visual Basic for Application

WMS – Web Map Service

ÚVOD

Světová města dlouhodobě ztrácejí status výrobních průmyslových center a stávají se centry terciárního a kvartérního sektoru. Počínaje 60. lety 20. století začínají tato města ztrácet většinu svých dělníků a zaměstnanců v průmyslu a to díky rozsáhlému uzavírání průmyslové výroby, mechanizaci, suburbanizaci průmyslu a obecně kvůli celosvětovému posunu industriální výroby z center do periferních částí světa. Město tvoří dynamický socioekonomický systém, který je vnitřně silně heterogenní. Tato heterogenita se zřetelně odráží také ve vnitřním prostorovém uspořádání města. Tyto proměny měly obrovský dopad právě na vnitřní strukturu města. Pokud tato industriální města chtěla hrát dále důležitou roli v národním či dokonce nadnárodním prostředí, musela projít celkovou proměnnou (Glasgow, Manchester, Lille, Bilbao, Dortmund, Katowice, Ostrava atd.). V tomto období ztrácela velkou část své populace, měnilo se ekonomické zaměření zaměstnanosti obyvatel a v neposlední řadě nastaly proměny v bytové výstavbě. Některé městské části tak prošly kompletní proměnou, stejně jako jejich obyvatelé. Urbanistické změny se rovněž promítají do změn v populaci, ať již jde o celkovou velikost, strukturu či jejich územní distribuci.

Z analytického hlediska lze změny v prostorové distribuci obyvatel či zaměstnanců sledovat prostřednictvím celé řady metrik. Popis distribuce, dokumentace nerovnoměrnosti či shlukování, sledování jejího vývoje a projevu může pomoci při sledování sociální situace ve městě, včas odhadovat a reagovat na negativní jevy, které prostorovou separaci a nerovnoměrnost doprovázejí.

Jako příklad českého industriálního města, které po revoluci prošlo značnou změnou, je v práci zpracováno území města Ostravy. Práce se zaměřuje na období 2009 – 2010, což je sice z pohledu sledování změn spojených s deindustrializací málo, avšak jen proto toto období jsou k dispozici dostatečně podrobná lokalizovaná data. Připravené postupy však bude možné využít pro další časová období pro hodnocení situace v jiných městech.

1 CÍLE PRÁCE

Cílem této diplomové práce je zhodnocení vybraných metod prostorové segregace a shlukování v městském prostředí na základě údajů o nezaměstnaných z Úřadu práce v Ostravě. Data z Úřadu práce v Ostravě byla vybrána, protože jsou lokalizovatelná až na adresní body.

Ze začátku je potřebné zjistit, jakých forem prostorová segregace nabývá a zhodnotit jí pomocí prostorových indikátorů (měření heterogenity územních celků). Jde o zjištění, zdali k nějaké výraznější segregaci či shlukování nedochází již na úrovni základního administrativního členění Ostravy.

V dalším kroku se zjišťuje prostorová autokorelace a testuje se vliv volby proměnných a nastavení parametrů zpracování, aby se zjistilo, které proměnné a jak moc jsou korelovány, aby se s nimi mohlo dále pracovat.

Dále se provede analýza LISA, která odhalí místa, kde dochází ke shlukování vysokých či nízkých hodnot sledovaných proměnných.

Kromě toho je rozebrána analýza shluků s následnou aplikací prostorového hierarchického shlukování.

Výsledky jsou porovnávány s předpokládanými lokalitami, které byly již dříve identifikovány experty.

2 FORMY A ASPEKTY PROSTOROVÉ SEGREGACE

Prostorová segregace (nedobrovolné vyloučení) je tím, co je v souvislosti se sociálním vyloučením zmiňováno nejčastěji. S tímto termínem se totiž pojí nejproblematičtější socioekonomické jevy jako existence ghett, slumů či jiných sídlišť osob na okraji společnosti.

Prostorové oddělení určitých skupin obyvatelstva se objevuje snad ve všech lidských kulturách s určitou minimální mírou společenské diferenciace. Různé společnosti rozdělovaly své členy v prostoru podle ekonomických, sociálních, etnických či náboženských hledisek. Např. již v „tradičních“ orientálních městech můžeme nalézt sdružování v prostoru na základě etnických či náboženských charakteristik. Ve středověkých evropských městech převažovalo oddělení určitých skupin obyvatelstva na základě sociální úrovně – kupecké čtvrti byly odděleny od čtvrtí řemeslníků, Židé žili v uzavřených ghettech. Rozdílnost městských čtvrtí včetně jejich převažujícího obydlení různými příjmovými a sociálními skupinami je jev, který k moderní urbánní společnosti patří.

Pojem segregace má dva odlišné významy – můžeme ho vnímat jako hodnotově neutrální termín, kdy segregací rozumíme prostě jen prostorové rozložení určitých skupin obyvatelstva, nebo ve smyslu normativním, podle kterého segregace představuje spíše společenský problém se závažnými společenskými důsledky.

2.1 Přístupy k hodnocení prostorové segregace

Temelová [36] uvádí, že k hodnocení prostorové diferenciace obyvatelstva na území měst a regionů lze přistupovat dvěma základními způsoby. První způsob představují etnografické studie a popis sociálního klimatu jednotlivých částí města zaměřený na jejich specifika v kontextu města jako celku. Důležitým zdrojem informací pro kvalitativní výzkum segregace jsou terénní výzkumy, rozhovory, pozorování, dotazníková šetření apod. Druhý přístup je založen na kvantitativním statistickém zhodnocení sociálně prostorové diferenciace, přičemž cílem je nalézání pravidelností, zobecňování zjištěných skutečností a vytváření modelů prostorového uspořádání. Sociální segregace představuje

vícerozměrný problém, kde všechny rozměry nejsou stejně významné. Obvykle jeden z nejdůležitějších faktorů je dlouhodobá nezaměstnanost. Hodnocení nezaměstnanosti s vhodným prostorovým rozlišením může být založeno též na údajích ze sčítání lidu, nebo ze záznamů z evidence nezaměstnaných osob. V posledním případě je důraz kladen na vhodné zpracování údajů z úřadů práce kvůli požadované ochraně soukromí. Horák [16] uvádí, že se podíl registrovaných nezaměstnaných na obyvatele v produktivním věku může nahradit míru nezaměstnanosti z důvodu vysoké korelace obou ukazatelů. Jeho další doporučené faktory pro sledování problémové situace na trhu práce jsou míra nezaměstnanosti mladých lidí (do 25 let), starších lidí (nad 50 let), podíl nezaměstnaných se základním vzděláním, podíl tělesně-postižených nezaměstnaných a podíl dlouhodobě nezaměstnaných.

Přestože podle Temelové [36] je určitý stupeň sociální a prostorové diferenciace společnosti přirozený, funkční a nevyhnutelný, existuje obecně sdílené přesvědčení, že vlády mohou zmírňovat výraznější nerovnosti v sociálně ekonomických podmínkách obyvatel různých rezidenčních prostředí a předcházet tak vzniku (příp. zmírňovat důsledky) rezidenční segregace. Politický zájem se zaměřuje především na problémy vyplývající z nedobrovolné segregace a vytváření koncentrací sociálně slabých, rasových a etnických menšin. Z toho však vyplývá určité nebezpečí v posilování nevraživosti a stigmatizace těchto skupin většinovou společností. Druhá strana sociálního spektra a vytváření „ghett bohatých“ jsou však často opomíjeny a nejsou vnímány jako problém, přestože k celkové úrovni sociálně-prostorového oddělení různých skupin obyvatel přispívají a ovlivňují tak naše vnímání společenské sounáležitosti a v důsledku i sociální soudržnost.

2.2 Mechanismy prostorového vyloučení

Proces globalizace vede ke zvýšené sociální polarizaci a následné ostřejší prostorové segregaci. Během posledních několika desetiletí došlo zejména na americkém kontinentu k značnému odlivu obyvatel z centrálních částí měst a tím se zvyšovala koncentrace chudoby. Jargowsky ve své práci [20] dokumentuje, že růst příměstských oblastí a pokles centrálních částí měst jsou projevy metropolitního vývojového procesu, který vede k vyšší úrovni hospodářské segregace. Rozšiřování měst a koncentrace chudoby jsou dvě strany

stejného metropolitního vývojového procesu a účinek tohoto procesu zvyšuje chudobu a limity rovnosti. Jargowsky ve své další práci [19] zjistil, že centra měst (na americkém kontinentu) okupují většinou chudé menšiny a na předměstí měst jsou bohatí obyvatelé. Existuje několik možností, jak ke vzniku vyloučených oblastí dochází. Prvním mechanismem vzniku vyloučených území je sestěhování vyloučených, které představuje krátkozraké řešení, ke kterým se mnohdy uchylují úřady obcí a měst, které se zabývají bytovou problematikou, ale ke kterým také mnohdy přispívá sociální politika svou tendencí koncentrovat chudé – sociální bydlení, rekvalifikační kurzy. Dochází tak k sestěhovávání nově příchozí chudiny v některých čtvrtích, kde se tak vytváří homogenní prostředí s mnoha nezamýšlenými důsledky.

Dobrovolná segregace vyjadřuje vůli členů jazykově nebo kulturně spřízněných skupin bydlet ve vzájemné blízkosti. Prostorová blízkost jim pak umožňuje lepší uspokojování společných potřeb a pomáhá jim uchovávat vlastní subkulturu.

Temelová ve své práci [36] uvádí, že závažnějším a naléhavějším společenským problémem je ale spíše segregace nedobrovolná. Jedná se o stav, kdy vyloučení v prostoru je vynucováno majoritní společností, podmínky života v lokalitě jsou zhoršené a její obyvatelé pociťují nespokojenost, chtěli by z lokality odejít, brání jim v tom ovšem finanční, institucionální, psychické bariéry apod.

2.3 Formy prostorové segregace

Rezidenční segregace nabývá mnoho forem. Liší se zejména v tom, zda jsou výrazem segregace dobrovolné nebo nedobrovolné a zda jsou charakteristické pro znevýhodněné a deprivované, nebo naopak pro privilegované.

2.3.1 Ghetto

Jak ve své práci [37] uvádí Toušek, představuje ghetto extrémní formu rezidenční segregace. Jedná se o oblast, ve které se nedobrovolně koncentrují členové určité sociální skupiny a mají v této oblasti převažující podíl. K nedobrovolné koncentraci dochází na základě sociálně-kulturních charakteristik, které mohou mít podobu etnicity, národnosti,

rasy, náboženství. Ghetta tedy mohou vznikat nezávisle na ekonomickém statusu, chudoba není základním určujícím faktorem ghetta.

Klasických ghett jsme byli často svědky i v minulosti. Typickým příkladem segregace na základě etnicity jsou z minulosti židovská ghetta. Po druhé světové válce ale přestává být ghetto domovem Židů, nově se stává domovem Afroameričanů, kteří byli segregováni do prostorově oddělených lokalit průmyslových měst na severu USA. Černošská ghetta zpočátku plnila zejména pozitivní funkci – stala se doslova kulturními centry, prosperujícími čtvrtěmi, později ovšem tato pozitivní funkce začala upadat a problematika byla postupně zbavena rasové konotace a změnila se spíše v oblast extrémní chudoby.

2.3.2 Enkláva

Na rozdíl od ghetta, se zde jedná o segregaci dobrovolnou. Marcuse ve své práci [26] definuje enklávu jako dobrovolně rozvinutou prostorovou koncentraci, ve které se soustřeďují příslušníci určité skupiny obyvatelstva za účelem zvýšení ekonomického, sociálního, politického nebo kulturního rozvoje vlastních členů. Enklávy mohou vznikat na základě kulturním, imigračním, nejčastěji se pak jedná o enklávy etnické. Základní rozdíl mezi ghetty a etnickými enklávami tedy spočívá v míře dobrovolnosti – ghetta vznikají odloučením nedobrovolným – segregací, etnické enklávy odloučením dobrovolným – separací. Enkláva může mít také, na rozdíl od ghetta, pozitivní význam.

2.3.3 Slum

Slum představuje chudinskou čtvrť, segregovanou, hustě osídlenou oblast, která se vyznačuje špatnou kvalitou bydlení a nedostatečnou úrovní základních služeb. Mezi další definiční znaky slumu patří absence nebo nízká kvalita inženýrských sítí, nedostatek pitné vody, špatná kvalita a minimální velikost obytných prostor, špatný zdravotní stav obyvatel, výskyt sociálně patologických jevů apod. S pojmem se poprvé setkáváme ve dvacátých letech devatenáctého století v Anglii, kde se užíval pro označení chudých dělnických čtvrtí v Londýně. Toušek ve své práci [37] uvádí, že pojem slum bývá často zaměňován s pojmem ghetto. Lidé ve slumech setrvávají nedobrovolně, stejně jako v ghettu, ale na

rozdíl od něj nevzniká segregace na základě etnicity, rasy, národnosti apod., jako v případě ghetta, ale na základě ekonomického statusu. Příčinou setrvávání lidí ve slumech je tedy primárně chudoba, tudíž je lidé opouštějí, jakmile svou špatnou finanční situaci překonají. Ghetta naopak představují pasti chudoby, v nichž jsou lidé trvale uvězněni a ze kterých je velmi nesnadné uniknout.

2.4 Důsledky segregace

Temelová [36] uvádí, že důsledky rezidenční segregace úzce souvisí s rozmanitými formami prostorových koncentrací různých skupin obyvatel. Pokud je prostorové oddělení vytvořeno na dobrovolné bázi, mohou být důsledky takového seskupení pro jeho obyvatele pozitivní. Podle Hájkové [13] prostorová blízkost lidí, kteří zastávají podobné sociální postavení, vyznávají podobné náboženství či patří ke stejnému etniku, přispívá ke zvýšení celkového pocitu bezpečí a sounáležitosti a také může vést k vytvoření hustých sociálních sítí. Segregace na dobrovolném základě také umožňuje rozvoj a uchovávání různých kultur, která by jinak, kdyby byli její příslušníci rozptýleni, mohla zaniknout. Někdy poskytuje koncentrované bydlení také ekonomické výhody, zejména pokud někteří jeho členové začnou rozvíjet různé ekonomické aktivity a poskytovat ostatním členům pracovní místa, ale také zboží a služby.

Mnohem více pozornosti si ovšem zaslouží segregace nedobrovolná, protože představuje závažný a naléhavý společenský problém, který má na obyvatele vyloučených lokalit řadu negativních důsledků. Segregace a izolace sociálně slabších a etnicky či rasově vymezených skupin obyvatel, ke které nedošlo z vlastní vůle jedinců, vytváří bariéry rozvoje životních šancí, protože cesta zpět je již velmi obtížná, mnohdy nemožná. Tyto bariéry mají podobu chátrajícího prostředí, bydlení v nedostatečně vybavených a poškozených bytech, které jsou často po povodních podmáčené, a díky tomu se v nich vyskytuje plíseň, což s sebou samozřejmě přináší rovněž velké hygienické problémy. Zdravotní a epidemiologické důsledky z bydlení v takto poškozených domech jsou šířící se infekční choroby jako například žloutenka typu A. Obvykle se jedná o oblasti s nedostatečně rozvinutými službami. Děti žijící v těchto oblastech navíc obvykle navštěvují školy špatné úrovně, což způsobuje neustálou reprodukci nerovností. Navíc Potter ve svém výzkumu [29] zjistil, že výše rodinných příjmů má značný vliv na výsledky

studentů. V těchto lokalitách se často objevuje deviantní chování, kriminalita, rozšířená je také konzumace drog a jiných návykových látek.

2.5 Prevence prostorové segregace

Mezi nejčastěji uváděné oblasti politických intervencí, které mohou přispět ke zmírnění důsledků či prevenci segregace, patří opatření v oblasti bytové politiky, územního plánování, sociální politiky a místního ekonomického rozvoje. Jak uvádí například Ostendorf v jeho výzkumu [28] a Temelová ve své práci [36], cílem politik městské restrukturalizace je diverzifikovat bytový fond čtvrtí a tím přispět ke zlepšení podmínek pro vzestupnou sociální mobilitu příjmově slabších obyvatel. Sociálně smíšených obytných zón je dosaženo restrukturalizací bytového fondu takovým způsobem, aby se v jedné čtvrti nacházely byty s různou kvalitou, cenou i způsobem užívání. Důležitým prostředkem je zlepšení kvality fyzického prostředí a služeb ve čtvrtích s vysokým podílem levného nájemního bydlení s cílem přilákat lidi vyšších příjmových kategorií.

Ne vždy je ale taková snaha korunována úspěchem. Některé zahraniční studie dokládají, že umělé budování smíšených čtvrtí vede k odchodu bohatší části obyvatel a opětovnému vzniku separace.

2.6 Měření prostorové segregace

S příchodem geografických informačních systémů máme mnoho možností, jak množství informací o segregaci zpracovávat. Při zpracovávání těchto dat se berou v úvahu podrobné informace o lokalizaci skupin obyvatelstva v městských oblastech, vlastnosti jednotlivých jednotek i vztahů se sousedními jednotkami. Prostorová data mají podle Spurné [34] mnoho specifických vlastností, které znesnadňují jejich analýzu a vyžadují použití odlišného souboru statistických metod, modelovacích přístupů i velmi citlivou interpretaci výsledků kvantitativních analýz. Standardní statistické metody vyvinuté pro analýzu neprostorových dat jsou tak v mnoha případech pro analýzu prostorových dat nevhodné. Za nejvýznamnější problémy či specifika analýzy prostorových považuje Spurná [34] závislost výsledků analýz na agregaci dat neboli na způsobu vymezení prostorových jednotek ve spojení s ekologickou chybou, prostorovou autokorelací a

prostorovou nestacionaritu. Zvláště u malých a členitých oblastí je potřeba podle Spurné [34] dále zmínit problémy vznikající v blízkosti hranice, která významným způsobem ovlivňuje výsledky statistických analýz. Uvedeným problémům je v oblasti prostorových analýz věnována největší pozornost a svým způsobem je jejich řešení impulsem pro další vývoj kvantitativních metod.

3 MĚŘENÍ HETEROGENITY ÚZEMNÍCH CELKŮ

K měření heterogenity se používají různé indikátory, zde označené jako indexy. Temelová a Sýkora ve své práci [36] jako nejpoužívanější indexy vyzdvihují index odlišnosti, index segregace, index interakce, index izolace a index koncentrace.

3.1 Index odlišnosti

Index odlišnosti/nepodobnosti (index of dissimilarity) je demografickým měřítkem pro hodnocení segregace mezi dvěma skupinami - [24]. Index tedy měří relativní rozdíl mezi prostorovým rozložením dvou jevů v rámci vnitřního územního členění města – [36]. Čím je hodnota indexu vyšší, tím více jsou skupiny odděleny (výsledky v intervalu <0,1>).

$$ID_{xy} = \frac{1}{2}\sum_{i=1}^{n}\left|\frac{x_i}{X} - \frac{y_i}{Y}\right|;\ 0 \le ID \le 1 \qquad (1)$$

kde: x_i, y_i = počet příslušníků skupin X a Y v i-té územní jednotce

X, Y = celkový počet příslušníků skupin X a Y v celém teritoriu (městě)

n = počet územních jednotek v daném teritoriu

Podle Martoriho a kolegů [24] se tento index jeví jako vhodný k použití v oblastech, kde jsou pouze dvě dominující skupiny. Naproti tomu v místech, kde existuje více skupin, není tento index tak spolehlivý.

3.2 Index segregace

Index segregace (index of segregation) udává relativní rozdíl mezi prostorovým rozložením jedné skupiny obyvatel a zbytkem populace. Index segregace vychází z indexu odlišnosti, který porovnává určitou skupinu obyvatel *X* s celkovou populací města *Y*, přičemž navíc zohledňuje velikost skupiny *X* (její podíl na celkové populaci) - [36]. Byl zaveden již v roce 1955 Duncanem a Duncanem - [24].

$$IS_{xn} = \frac{ID_{xn}}{\left[1 - \left(\frac{X}{Y}\right)\right]};\ 0 \le IS \le 1 \qquad (2)$$

kde: ID_{xn} = index odlišnosti mezi skupinou *X* a celkovou populací *Y*

X = celkový počet příslušníků skupiny *X* v daném teritoriu

Y = celkový počet obyvatel v daném teritoriu

Index se pohybuje v rozmezí od 0 (daná oblast je bez segregace) do 1 (maximální segregace). Výsledek může být vyjádřen v procentech a uvádí podíl členů menšinové skupiny, která musí změnit své bydliště v oblasti, aby se dosáhlo rovnoměrného rozdělení. Tato koncepce interpretace pomocí procent byla ovšem kritizována, protože nebere v úvahu nově získaný prostor, avšak nám dává velmi praktickou představu o výši nerovnoměrného rozdělení - [24].

3.3 Index izolace

Index izolace (index of isolation) je založen na potenciálním kontaktu mezi jedinci žijícími v jedné oblasti a jejich situaci v izolaci. Počítá pravděpodobnost interakce člena menšinové skupiny s jiným členem stejné skupiny - [24] [36].

$$II = \sum_{i=1}^{n} \left(\frac{x_i}{X} \right) \cdot \left(\frac{x_i}{t_i} \right);\ 0 \le II \le 1 \qquad (3)$$

kde: x_i, = počet příslušníků skupiny X v i-té územní jednotce

t_i= celkový počet obyvatel v i-té územní jednotce

X = celkový počet příslušníků skupiny X v celém teritoriu (městě)

n = počet územních jednotek v daném teritoriu

3.4 Index interakce

Expoziční index (exposure index), někdy také index interakce, je založen na potenciálním kontaktu mezi jedinci žijícími v jedné oblasti. Vyjadřuje pravděpodobnost toho, že jeden člen menšinové skupiny spolupracuje s členem referenční skupiny - [24], [36].

$$EI = \sum_{i=1}^{n} \left(\frac{x_i}{X} \right) \cdot \left(\frac{y_i}{t_i} \right);\ 0 \le EI \le 1 \qquad (4)$$

kde: x_i, y_i = počet příslušníků skupin X a Y v i-té územní jednotce

t_i = celkový počet obyvatel v i-té územní jednotce

X = celkový počet příslušníků skupiny X v celém teritoriu (městě)

n = počet územních jednotek v daném teritoriu

3.5 Index koncentrace

Index koncentrace (Concentration of poverty (affluence)) – základní myšlenka tohoto indexu je, že menšinové skupiny obvykle žijí v hustěji obydlených oblastech, vzhledem k jejich nižším ekonomickým možnostem. Výsledek se opět pohybuje od 0 do 1, kde 0 znamená, že segregace není - [24].

$$COP = \frac{1}{2}\sum_{i=1}^{n}\left|\frac{x_i}{X} - \frac{a_i}{A}\right|;\ 0 \le COP \le 1 \qquad (5)$$

x_i = počet příslušníků skupiny X v i-té územní jednotce

X = celkový počet příslušníků skupiny X v celém teritoriu (městě)

a_i = plocha i-té jednotky

A = celková rozloha města

n = počet územních jednotek v daném teritoriu

4 PROSTOROVÁ AUTOKORELACE

Podle Nezdařilové [27] z prostého překladu pojmu prostorová autokorelace vyplývá nejlépe jeho obsah, jímž je korelace jednoho jevu se sebou samým v prostoru, která se projevuje statisticky významným uspořádáním hodnot sledovaného jevu v prostoru nebo jak uvádí Horák ve své práci [17], že se obecně posuzuje podobnost charakteristik sledovaných objektů v závislosti na jejich vzdálenosti. Jinou definici uvádí Anselin [3] a to takovou, že princip prostorové autokorelace lze v nejobecnějším pohledu chápat jako existenci určitého funkčního vztahu mezi pravděpodobností výskytu určitého jevu v prostorové jednotce *i* a pravděpodobnosti výskytu tohoto jevu v jednotkách *j*, které jsou jí prostorově blízké. Formálně ho lze tedy vyjádřit ve tvaru:

$$p_i(y) = f\left(\sum_i w_{ij} p_j(y)\right), \qquad (6)$$

kde $p_i(y)$ je pravděpodobnost výskytu jevu y v jednotce i, w_{ij} pro $i \neq j$ jsou váhy. Pokud analyzovaná data vykazují pozitivní prostorovou autokorelaci, vytváří zároveň shluky jednotek s podobnými hodnotami sledovaného jevu. Naopak, pokud vysoké hodnoty tíhnou k tomu nacházet se v těsné blízkosti nízkým hodnotám a naopak, jedná se o negativní prostorovou autokorelaci. Pokud jsou data lokalizována tak, že neexistuje žádný vztah mezi blízkými hodnotami, hovoříme o nulové prostorové autokorelaci.

Stejně jako v případě párové korelace neukazuje signifikantní prostorová autokorelace na kauzální vztah a nevypovídá nic o příčině sledovaného uspořádání, proto je důležité pečlivé zkoumání a rozbor situace.

4.1 Měření prostorové autokorelace

Podle Anselina [3] lze prostorovou autokorelaci měřit několika odlišnými prostorovými autokorelačními statistikami popisujícími podobnost blízkých pozorování v závislosti na skutečnosti, že se jedná o diskrétní či spojitou proměnnou. Obecně každá statistika prostorové autokorelace dává do souvislosti atributovou podobnost c_{ij} a vzdálenost (blízkost) w_{ij} prostorových jednotek i a j v nejjednodušším vyjádření:

$$\sum_i \sum_j c_{ij} w_{ij} . \qquad (7)$$

4.1.1 Moranovo I kritérium

Spurná ve své práci [34] uvádí, že všechny autokorelační statistiky závisí na nějaké definici prostorového vážení (u semivariogramů a autokorelačních funkcí se nic jiného než prostá vzdálenost nepoužívá), která se pokouší kvantifikovat často subjektivní koncepty blízkosti, a vzájemně se liší vyjádřením atributové podobnosti c_{ij}. V současnosti je jedním z nejpoužívanějších ukazatelů sloužících k měření prostorové autokorelace Moranovo I kritérium, které je definováno vzorcem:

$$I = \frac{n \cdot \sum_i \sum_j w_{ij} c_{ij}}{s^2 \sum_i \sum_j w_{ij}}, \qquad (8)$$

kde: $c_{ij} = (z_i - \bar{z})(z_j - \bar{z})$ a $s^2 = \sum_i (z_i - \bar{z})^2$,

přičemž n je počet analyzovaných jednotek, i, j jsou indexy charakterizující nějaké dvě jednotky, z_i značí hodnotu proměnné v jednotce i a z aritmetický průměr sledované proměnné. Proměnná vykazuje pozitivní prostorovou autokorelaci, pokud je hodnota

Moranova I kritéria kladná, a negativní prostorovou autokorelaci, pokud je hodnota Moranova I kritéria záporná. Hodnoty Moranova I kritéria blízké nule poukazují na nulovou prostorovou autokorelaci.

4.1.2 LISA

Podle Spurné [34] jsou v současné době nejvíce používány lokální indikátory prostorové asociace (local indicators of spatial association – LISA) vyvinuté Anselinem, které se staly standardním nástrojem pro lokální analýzu prostorové autokorelace.

Anselin ve své práci [1] uvádí, že zvýšená dostupnost velkých sad prostorových dat a sofistikovaných funkcí pro vizualizaci, rychlé vyhledávání dat a manipulaci v geografických informačních systémech, vznikla poptávka po nových technikách prostorové analýzy dat. Ačkoliv mnoho metod je k dispozici již v panelech nástrojů geografických analýz, jen málo z nich je vhodných k zabývání se explicitně prostorovými aspekty.

Jak podotkla Spurná ve své práci [34], analýza LISA úzce souvisí s Moranovým diagramem, prostřednictvím kterého lze vykreslit primární závěry analýzy prostorové autokorelace. „V tomto diagramu s původními hodnotami proměnné na horizontální ose a vypočítanými průměrnými hodnotami ze sousedních jednotek na vertikální ose odpovídá sklon proložené regresní přímky hodnotě Moranova I kritéria. Se záměrem ulehčení interpretace jsou přitom analyzované proměnné standardizovány. Na základě výpočtu LISA můžeme uskutečnit kategorizaci sledovaných jednotek v souladu s typem prostorové autokorelace do čtyř skupin, které odpovídají čtyřem kvadrantům v Moranově diagramu. Prostorové shluky, které vykazují nadprůměrné či podprůměrné hodnoty proměnné v určité jednotce souhlasně s jejím okolím, se v grafu nalézají v pravém horním (hot spots, hodnota vysoká-vysoká) a levém dolním (cold spots, hodnota nízká-nízká) kvadrantu. Případné prostorové odchylky (spatial outliers) typické nadprůměrnou / podprůměrnou hodnotou proměnné v určité jednotce a podprůměrnými / nadprůměrnými hodnotami v jejím okolí v pravém dolním (hodnota vysoká-nízká) / levém horním (hodnota nízká-vysoká) kvadrantu. Výsledky analýzy LISA mohou být vizualizovány v mapové podobě, přičemž vyobrazit lze jak statistickou významnost charakteristiky LISA pro jednotlivé územní jednotky, tak samozřejmě výše uvedené řazení jednotek do kategorií se signifikantními

hodnotami“ [34]. Aktivum statistické analýzy LISA je podle Spurné [34] ostřejší zobrazení oblasti s nadprůměrnými nebo opačně s podprůměrnými hodnotami monitorovaného ukazatele, než umožňuje metoda kartogramu, která je pouhým vizualizačním nástrojem, včetně statistického vyhodnocení tvorby prostorových shluků. Při podrobnějším studiu je důležitá samotná identifikace územních jednotek, které se hodnotou ukazatele nápadně odlišují od svého okolí. Právě podrobnější studium původu existence prostorových shluků a naopak jednotek odlišujících se od svého okolí by mohlo být v mnohém přínosné. Z porovnání s číselným výsledkem globální analýzy prostorové autokorelace je nepopíratelné, že analýza LISA dává mnohem důkladnější a přesnější informace o povaze prostorové autokorelace v rámci sledovaného území a je v podstatě nepostradatelným doplňkem globální analýzy.

5 ANALÝZA SHLUKŮ

Podle Melouna a kol. [25] patří analýza shluků (Cluster analysis, CLU) mezi metody, které se zabývají vyšetřováním podobnosti vícerozměrných objektů a jejich klasifikací do tříd čili shluků. Hodí se zejména tam, kde objekty projevují přirozenou tendenci se seskupovat. Ve své podstatě mezi hlavní tři cíle analýzy shluků patří:

- popis systematiky, což je tradiční využití shlukové analýzy pro průzkumové cíle a taxonomii, což je klasifikace objektů, která je založená na zkušenosti,
- zjednodušení dat, kdy nám vlastní shluková analýza poskytne při hledání klasifikace poznávacích cílů zjednodušený pohled na objekty,
- identifikaci vztahu, kde poté po samotném nalezení shluků sledovaných objektů a tím i přesné struktury mezi jednotlivými objekty, je pro nás daleko snadnější odhalit meziobjektové vztahy.

Cíle analýzy shluků nelze odloučit od hledání a výběru vhodných znaků k typizování shlukovaných objektů. Objevené shluky vyjadřují skladbu dat pouze se zřetelem na vybrané znaky. Volba znaků musí být uskutečněna podle teoretických, pojmových a realistických hledisek. Vlastní analýza shluků nezahrnuje techniku k rozlišení podstatných a nepodstatných znaků. Uskuteční výhradně diferenciaci shluků. Špatné zahrnutí znaků vede k zařazení i odlehlých objektů, které mohou rušivě ovlivnit závěry analýzy. Měly by být užity výhradně takové znaky, jenž uspokojivě dělají rozdíly mezi objekty.

5.1 Míry podobnosti

Jak uvádí Meloun a kol. ve své práci [25], myšlenka podobnosti objektů je v analýze shluků základní. Podobnost mezi objekty je uplatněna jako kritérium produkce shluků objektů. Zpočátku se určí znaky stanovující podobnost, které se dále sdružují do podobnostních měr. Takto potom může být objekt srovnán s dalším objektem. Analýza shluků tvoří shluky obdobných objektů. Meziobjektová analogie může být měřena rozmanitými prostředky, které se dají zpravidla zahrnout do jedné ze tří elementárních kategorií a to míry korelace, míry vzdálenosti a míry asociace. Každá z nich představuje

zvláštní pohled na podobnost, která je závislá na objektech a na typu dat. Korelační a vzdálenostní míry jsou míry metrických dat, zatímco asociační míry jsou určeny spíše pro nemetrická data.

5.1.1 Korelační míry

Podle Melouna a kol. [25] může být primární mírou podobnosti dvou objektů či vlastností x_i a x_j, formulovaných v kardinální škále Pearsonův párový korelační koeficient r. Objekty jsou si tím podobnější, čím je jejich párový korelační koeficient větší a bližší jedničce. V případě ordinální škály (pořadová čísla) je obdobnou mírou podobnosti Spearmanův korelační koeficient. Obyčejně se vychází z transponované matice dat X^T, kdy sloupce představují objekty a řádky pak znaky. Korelační koeficienty mezi dvěma sloupci matice X^T představují korelaci mezi dvojicí objektů. Tomu odpovídá podobnost jejich profilů v profilovém diagramu. Vysoká korelace indikuje vysokou "podobnost" a nízká korelace pak "nepodobnost" profilů.

5.1.2 Míry vzdálenosti

Podle Melouna a kol. [25] představují míry vzdálenosti nejčastěji používané míry, založené na prezentaci vzdálenosti mezi objekty v prostoru, jehož souřadnice tvoří jednotlivé znaky. Nejobvyklejší vzdálenostní mírou je Euklidovská vzdálenost známá také jako geometrická metrika, která představuje délku přepony pravoúhlého trojúhelníka a vypostí se pomocí Pythagorovy věty. Platí, že vzdálenost

$$d_E(x_k, x_l) = \sqrt{\sum_{j=1}^{m} \left(x_{k,j} - x_{lj}\right)^2} \qquad (9)$$

reprezentuje standardní typ vzdálenosti. Vedle Euklidovské vzdálenosti se používá také čtverec Euklidovské vzdálenosti, který představuje základ Wardovy metody shlukování.

Mnohdy používaná Manhattanská vzdálenost je označovaná také jako vzdálenost městských bloků nebo jako Hammingova metrika, formulovaná vztahem

$$d_H(x_k, x_l) = \sum_{j=1}^{m} \left| x_{kj} - x_{lj} \right| \qquad (10)$$

Před použitím této vzdálenosti se musíme přesvědčit, že znaky spolu nekorelují. Jestliže tento předpoklad není splněn, shluky jsou nesprávné. Následující míra je zobecněná Minkovského metrika, pro kterou platí

$$d_M(x_k, x_l) = \sqrt[z]{\sum_{j=1}^{m} \left| x_{kj} - x_{lj} \right|^z} \qquad (11)$$

kde pro *z = 1* jde o Hammingovu metriku a pro *z = 2* o Euklidovu. Čím dosahuje z vyšších hodnot, tím více je vyzdvihována odchylka mezi vzdálenými objekty. V některých situacích se užívá rovněž tětivová vzdálenost (chord distance), definovaná vztahem

$$d_{CH}(x_k, x_l) = \sqrt{2\left[1 - \frac{\sum_{j=1}^{m} x_{kj} x_{lj}}{\sum_{j=1}^{m} x^2{}_{kj} \sum_{j=1}^{m} x^2{}_{lj}}\right]} \qquad (12)$$

V případě, že se použijí tři znaky, je tětivová vzdálenost přímou vzdáleností dvou bodů na povrchu koule s jednotkovým poloměrem a počátkem v těžišti.

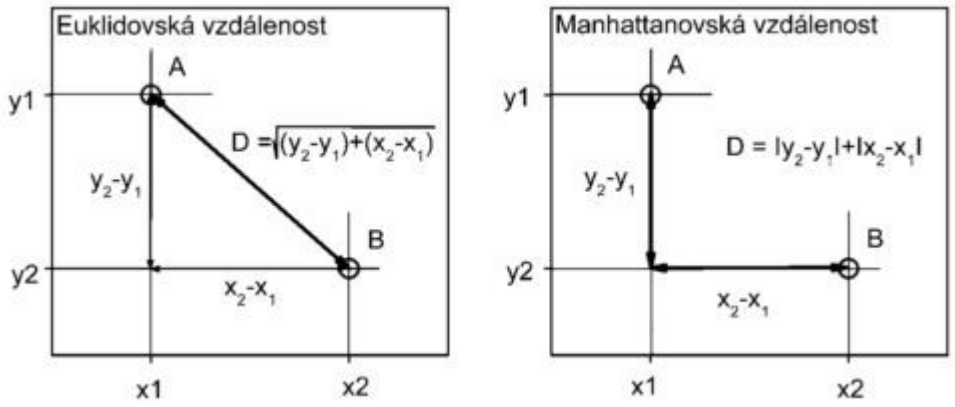
Potíž veškerých vzdálenostních měr vznikne, když použijeme nestandardizovaná data, která mohou zapříčinit odlišnosti mezi shluky, zásluhou mnohdy veliké odlišnosti jednotek měření. Shluky rozličných vzdálenostních měr se budou odlišovat, největší

rozptýlenost mezi shluky bude u čtverce Euklidovské vzdálenosti. Posloupnost podobností se podstatně obmění se změnou měřítka nebo proměnou jednotek jednoho ze znaků.

Všechny dosud uvažované metriky neuvažují závislost mezi znaky. Zařadíme-li do vztahu pro vzdálenost rovněž vazby mezi znaky, vyjádřené kovarianční maticí *C*, obdržíme novou statistickou míru, zvanou Mahalanobisova metrika

$$d_{Ma}(x_k, x_l) = \sqrt{(x_k - x_l)^T C^{-1} (x_k - x_l)} \qquad (13)$$

Ve skutečnosti jde o vzdálenost bodů v prostoru, jehož osy nemusí být ortogonální. Vysoce korelovaná selekce znaků může skrytě převážit celý soubor znaků shlukování.



Obr. 1.: Nejpoužívanější míry vzdálenosti: (a) Euklidovská D, (b) Manhattanovská vzdálenost D (převzato z Melouna a kol. [25])

5.1.3 Míry asociace

Míry asociace podobnosti se užívají ke srovnání objektů, jak uvádí Melou a kol. ve své práci [25], pokud jsou jejich znaky nemetrického charakteru (například binární proměnné). Mezi základní koeficienty podobnosti patří Sokalův-Michenerův koeficient asociace (zvaný také koeficient jednoduché shody), Russelův-Raoův koeficient asociace, Jaccardův

koeficient, Hamannův koeficient asociace, Korelační koeficient, Rogersův a Tanimotův koeficient asociace a Sörensenův koeficient asociace.

5.2 Standardizace dat

Před provedením vlastní shlukové analýzy je třeba vyřešit ¨problém, zda je třeba data standardizovat. Musí se respektovat skutečnost, že většina měr vzdáleností je velmi senzitivních na volbu měřítka (stupnice), která, jak uvádí Meloun a kol. [25], vede k různým numerickým velikostem znaků. Obecně je v platnosti pravidlo, že znaky s větší mírou proměnlivosti čili větší směrodatnou odchylkou mají větší dopad na míru podobnosti.

5.3 Způsoby shlukování

Meloun a kol. [25] ve své práci uvádí, že shluk (cluster) je skupina objektů, jejichž vzdálenost (nepodobnost) je menší než vzdálenost, resp. nepodobnost, kterou mají objekty, které do shluku nenáleží. Na základě způsobu shlukování se postupy dělí na hierarchické shlukování a nehierarchické shlukování. Hierarchické se rozdělují dále na aglomerační shlukování a divizní shlukování.

5.3.1 Hierarchické shlukovací postupy

Podle Melouna a kol. [25] se zakládají na hierarchickém uspořádání objektů a jejich shluků. Graficky se hierarchicky vytvořené shluky zobrazují pomocí vývojového stromu nebo dendrogramu. U aglomeračního shlukování se do prvního shluku spojí dva objekty, jejichž vzdálenost je nejmenší a vypočte se nová matice vzdáleností, v níž jsou vynechány objekty z prvního shluku. Tento shluk je poté zařazen jako objekt. Tento postup se opakuje do té doby, dokud veškeré objekty nepředstavují jeden velký shluk nebo dokud nezbude předem zadaný počet shluků. Proces divizního shlukování je opačný. Vychází se z množiny veškerých objektů jako jediného shluku a jeho, po sobě jdoucím dělením. Tím získáme systém shluků, až nakonec skončíme ve stadiu jednotlivých objektů. Výhodou hierarchických metod je nepotřebnost informace o optimálním počtu shluků v procesu

shlukování; tento počet se určuje až dodatečně. Při shlukování vznikají pouze dva základní problémy, prvním je způsob vyjádření podobnosti mezi objekty a druhým je volba vhodné shlukovací procedury. Volba vhodné shlukovací procedury souvisí se zvoleným způsobem vyjádření metriky. Mezi metody metriky shlukování patří:

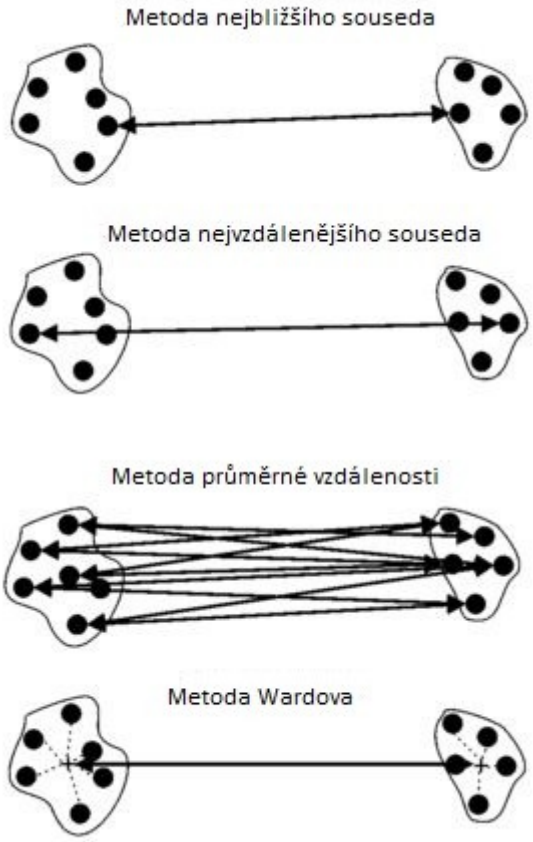
1. Metoda nejbližšího souseda. Metoda je postavena na minimální vzdálenosti. Podle Řezankové [32] se v prvním kroku naleznou dva objekty, které jsou oddělené nejkratší vzdáleností a umístí se do prvního shluku. Následující shluk vznikne přidáním třetího nejbližšího objektu a to tak, že se najde minimální vzdálenost mezi objekty. Proces se opakuje do té doby, než jsou všechny objekty v jediném společném shluku. Častou nevýhodou metody nejbližšího souseda je řetězový efekt, kdy se spojují shluky, jejichž dva objekty jsou sice nejbližší, ale vzhledem k většině ostatních objektů nejde o nejbližší shluky.

2. Metoda nejvzdálenějšího souseda. Jde o metodu podobnou předchozí kromě toho, že kritérium je vybudováno ne na minimální, ale na maximální vzdálenosti vektoru mezi všemi páry proměnné získaná z dvou shluků. Nejdelší vzdálenost mezi objekty v každém shluku si můžeme představit jako nejmenší kouli, která obklopuje všechny objekty v obou shlucích. Tato metoda se také někdy pojmenovává jako metoda úplného propojení, poněvadž všechny objekty ve shluku jsou propojeny každý s každým při maximální vzdálenosti tedy minimální podobnosti. Může se říci, že podobnost uvnitř shluku je rovna průměru shluku. Obě míry vystihují pouze jedno hledisko. Nejkratší vzdálenost postihuje pouze jednoduchý pár nejtěsnějších objektů a nejvzdálenější postihuje také jediný pár, ale pár dvou extrémů. Tato metoda je silně zaujatá vůči výrobě kompaktních shluků s podobnými průměry a může být vážně narušena odlehlými hodnotami. Je to metoda, která zaručuje, že všechny položky ve shluku mají minimální vzdálenost od jiných.

3. Metoda průměrné vzdálenosti. Kritériem vytvoření shluků je průměrná vzdálenost veškerých objektů v jednom shluku ke všem objektům ve shluku druhém. Takové techniky nejsou ovlivněny extrémními hodnotami, jako je tomu u metody nejbližšího souseda nebo u metody nejvzdálenějšího souseda, ale vznik shluku

záleží na všech objektech shluku, a ne jenom na jediném páru dvou extrémních objektů.

4. Wardova metoda. Principem této metody není klasická optimalizace vzdáleností mezi shluky ale minimalizace heterogenity shluků dle kritéria minimálního přírůstku vnitroskupinového součtu čtverců odchylek objektů od těžiště shluků. V jednotlivých krocích se pro všechny dvojice odchylek vypočítá nárůst součtu čtverců odchylek, který vznikne jejich sloučením a pak se sjednotí ty shluky, které obsahují minimální hodnotu tohoto přírůstku.
5. Metoda těžiště. U této metody se jedná o vzdálenost dvou těžišť shluků reprezentovaných eukleidovskou vzdáleností nebo čtvercem eukleidovské vzdálenosti. Těžiště shluku má souřadnice, které odpovídají průměrným hodnotám objektů pro jednotlivé znaky. Po každém kroku shlukování se vypočítává nové těžiště. Poloha těžiště shluku se stěhuje tak, jak se přičleňují nové objekty a vytváří se větší shluky. Mohou se objevit také zmatečné shluky, když vzdálenost mezi těžišti jednoho páru je menší než vzdálenost mezi těžišti jiného páru utvořeného v předešlém kroku. Výhodou této metody je menší ovlivnění odlehlými body, než je tomu u ostatních hierarchických metod.
6. Metoda mediánová. Jde o jakési vylepšení metody těžiště, neboť usiluje o odstranění rozdílné významnosti, které metoda těžiště dává rozmanitě velkým shlukům.



Obr. 2.: Nejčastěji užívané metriky shlukování (upraveno dle Melouna a kol [25])

5.3.2 Nehierarchické shlukovací postupy

K těmto postupům patří podle Melouna a kol. [25] metoda zárodečných bodů (Seeded). Uživatel podle jeho věcných znalostí stanoví, jaké objekty mají utvořit zárodky nově vytvořených shluků, a metoda pak rozčlení objekty do shluků podle jejich eukleidovské vzdálenosti od těchto charakteristických objektů.

5.3.3 Shlukování metodou nejbližších těžišť (K-Means)

Meloun a kol. [25] uvádí, že metoda nejbližších těžišť poskytuje pouze jediné řešení pro zadaný počet požadovaných shluků. Počet shluků musí být předem zadán uživatelem. Postup je založen na nejbližším těžišti, kdy je objekt zařazen do shluku s nejmenší vzdáleností mezi objektem a těžištěm shluku. Konkrétní technika zařazení objektu závisí

na dostupné informaci. Jsou-li těžiště shluků známá, mohou být specifikována v datech a zařazení objektu je založeno na nich. Jinak jsou těžiště shluků určována iteračním výpočtem z dat.

Princip metody nejbližších těžišť (K-means) spočívá v rozdělení *n* objektů o *m* znacích do k shluků tak, že mezishluková suma čtverců je minimalizována. Jelikož počet možných uspořádání je enormně veliký, nelze očekávat vždy jediné a nejlepší řešení. Algoritmus nalezne vždy spíše optimum lokální než globální. Jde o takové uspořádání shluků, kdy přemístění objektu z jednoho shluku do druhého nezpůsobí snížení sumy čtverců. Algoritmus pracuje iterativně, startuje vždy z jiného počátečního uspořádání. Nakonec vybere vhodné řešení ze všech možných dosažených uspořádání shluků.

5.3.4 Shlukování metodou optimálních středů čili medoidů

Medoid, čili optimální střed shluku, je podle Melouna a kol. [25] takový střední objekt, pro který platí, že průměrná vzdálenost k ostatním objektům v tomto shluku je minimální. Je-li požadováno k shluků, bude existovat také k medoidů. Po nalezení medoidů jsou data klasifikována do shluků vždy okolo nejbližšího medoidu. Medoidy a shluky se vytvářejí na základě vzdáleností čili nepodobností.

1. Späthova metoda: Meloun a kol. [25] uvádí, že tato metoda minimalizuje účelovou funkci přemísťováním objektů z jednoho shluku do druhého. Začíná u počátečního uspořádání shluků, algoritmus pak najde lokální minimum inteligentním přesouváním objektů ze shluků do shluku. Jakmile se nepřemístí už žádný objekt, metoda končí proces. Lokální minimum však nemusí být globálním. Aby program překonal toto omezení, zopakuje se několikrát hledání vždy z jiného startovacího uspořádání a nejlepší uspořádání shluků je nakonec bráno za výsledné.

Jako účelová funkce se bere celková vzdálenost mezi všemi objekty ve shlucích podle vzorce

$$D = \sum_{l=1}^{k} \sum_{l \in c_k} \sum_{j \in c_k} d_{ij} \,, \tag{14}$$

kde k je celkový počet shluků, d_{ij} představuje vzdálenost mezi i-tým a j-tým objektem a c_k udává soubor všech objektů ve shluku l.

2. Metoda PAM (Partition Around Medoids): Podle Melouna a kol. [25] minimalizuje celkovou vzdálenost D, proces postupuje takto:
    1. Nalezne se reprezentativní soubor k objektů. První objekt má nejkratší vzdálenost ke všem ostatním objektům, čili představuje střed shluku, medoid. Pak se $k - 1$ objektů hledá tak, že hodnota D je co možná nejmenší.
    2. Možné alternativy polohy k objektů jsou vybírány iteračním způsobem. Algoritmus vyhledává dosud nezařazené objekty a přemísťuje je tak, aby se hodnota D snižovala. Iterace skončí, jakmile změny nezpůsobí další snížení hodnoty D.
3. Silueta: Meloun a kol. [25] uvádí, že jde o statistické kritérium, které poskytuje klíčovou informaci o dobrém a špatném shluku. Hodnota siluety s se vypočte tímto způsobem:
    1. Objekt i je ve shluku A a má průměrnou vzdálenost a ke všem objektům ve svém shluku. Je-li ve shluku A jediný objekt, je $a = 0$.
    2. Sousední shluk B obsahuje objekty, které jsou nejbližší k objektu i ve shluku A a b je průměrná vzdálenost mezi objektem i a všemi objekty ve shluku B.
    3. Silueta s objektu i se vyčíslí tak, že pokud shluk A obsahuje pouze jeden objekt, je $s = 0$. Když $a < b$, je $s = 1 - a/b$. Když $a > b$, je $s = b/a - 1$. Když $a = b$, *je* $s = 0$.

Silueta se vyčíslí pro každý objekt. Hodnota siluety se mění od –1 do +1 a je mírou úspěšné klasifikace do shluků při porovnání vzdáleností uvnitř shluku.

5.4 Postup analýzy shluků

Jak uvádí Meloun a kol. ve své práci [25], analýza shluků poskytuje uživateli empirické a objektivní metody k provádění jedné z nejzákladnějších činností člověka – klasifikaci. Analýza shluků je pokaždé účinným analytickým nástrojem k účelům zjednodušení, rekognoskace a potvrzení, který má rozsáhlou oblast použití. Tato analýza má bohužel ještě řadu bílých míst, které přinesou často i zkušenému uživateli problémy v rozhodování. Když se analýza shluků užije správně, může odhalit strukturu v datech, kterou by jinak nešlo nalézt. Postup analýzy shluků krok za krokem:

1. Zvolíme vstupní databázi: zadává se typ dat (a) proměnných (sloupců) analyzovaných objektů (řádků), (b) sloupců matice vzdáleností, (c) sloupců korelační matice.
2. Volba druhu veličin: zadává se typ užitých veličin v datech, která mohou být (a) intervalová, (b) ordinální, (c) nominální, (d) symetrická binární, (e) asymetrická binární, (f) poměrová.
3. Zadání názvu objektů: zadáme pojmenování či jména jednotlivých objektů, umístěných v řádcích, které se mohou objevit v dendrogramu místo indexů (pořadových čísel) objektů.
4. Vybereme typ shlukovací techniky: volba metody z možností: jednoduchá průměrová (Average), skupinového průměru, centroidní (Centroid), nejbližšího souseda (Single, Nearest), nejvzdálenějšího souseda (Complete, Furthest), mediánová (Median), Wardova, a flexibilní.
5. Zvolíme druh užité vzdálenosti: vzdálenosti mohou být Euklidova metrika čili geometrická vzdálenost, Hammingova metrika čili Manhattanská vzdálenost, zobecněná Minkowskiho metrika a Mahalanobisova metrika.
6. Postup linkování a zařazení do shluků: tabelární výpočet vzdáleností (nebo podobností) mezi objekty a shluky a postupné vytváření dendrogramu. Postupy

jsou (1) metodou hierarchického shlukování, (2) shlukování metodou nejbližších středů, (3) shlukování metodou středů-medoidů, a (4) metodou fuzzy shlukování.

7. Vypočteme skutečné a predikované vzdálenosti v dendrogramu: porovnáme skutečné vzdálenosti mezi objekty a vypočtené vzdálenosti (predikované) v dendrogramu, jejich rozdíl a konečně i procentuální vyjádření tohoto rozdílu.
8. Hledáme nejlepší techniky tvorby dendrogramu: dle bodu 4. a 5. lze k sestrojení optimálního dendrogramu kombinovat řadu technik. Rozhodčím kritériem věrohodnosti jsou především kofenetický korelační koeficient *CC*, obě míry těsnosti proložení delta, ale také další kritéria: mezishluková suma čtverců WSS_K, procento variace PV_K, silueta *s*, průměrná silueta *SC*, Wilkova statistika λ, rozdělovací koeficienty Dunnův *F(U)* a Kaufmanův *K D(U)*.
9. Vysvětlení nejlepšího dendrogramu podobností objektů: interpretace optimálního dendrogramu podobnosti jednotlivých objektů je prvním a nejdůležitějším cílem shlukové analýzy.

Meloun a kol. ve své práci [25] uvádí, že dosažení konečného počtu shluků je snad nejvíce klamnou otázkou v analýze shluků, někdy též označovaná jako terminační kritérium. Neexistuje ani jeden objektivní prostředek stanovení tohoto kritéria. Jedno z terminačních kritérií se týká poměrně elementárního vyšetření měr podobnosti mezi shluky v každém kroku. Pokud totiž míra podobnosti přesáhne předdefinovaný rozměr, nebo když se následující hodnoty skokově změní, je vhodné postupovat tak, že si určíme různý počet shluků např. 2, 3 a 4 a podle uvažování o alternativním výsledku, realistickém mínění a teoretických základech úlohy samé se rozhodne. Jestliže se nalezne jednoobjektový shluk nebo shluk o poměrně malém rozsahu, uživatel se musí sám rozhodnout, jestli tento reprezentuje konstruktivního příslušníka vzorku nebo zda ho lze prohlásit jako nedostatečně reprezentativní pro soubor dat.

5.4.1 Cíle analýzy shluků

Podle Melouna a kol. [25] je primárním cílem analýzy shluků rozčlenění souboru objektů do dvou nebo více skupin, tříd či shluků, založených na podobnosti objektů, a to

dle předem vytipovaných znaků. Při vytváření homogenních shluků objektů se soustředíme na tři cíle:

a) Popis systematiky. Klasickým použitím shlukové analýzy jsou průzkumové cíle a popis systematiky – taxonomie, tj. empirická klasifikace objektů. Analýzou shluků lze generovat hypotézy spojené se strukturou objektů. Přestože se analýzy shluků používá v první řadě v průzkumové analýze, lze ji užít též v konfirmatorní analýze. Shlukovou analýzou se dojde k určitým shlukům objektů, které se dále srovnávají s jejich teoreticky odvozenou typologií.

b) Zjednodušení dat. Během hledání taxonomie poskytuje analýza shluků zjednodušený pohled na objekty. Objevené shluky objektů jsou připraveny k další následné analýze. Zatímco faktorová analýza usiluje o nalezení stavby znaků, analýza shluků koná totéž, avšak pro objekty. Na objekty se dále už nehledí jako na jeden jednotný soubor, ale jako na odloučené shluky objektů, rozdělené podle jejich vlastností.

c) Identifikaci vztahu. Po nálezu shluků objektů, a tím i vazeb mezi objekty je jednodušší objevit vztahy mezi objekty, což by bylo mezi samotnými objekty mnohem obtížnější. Shluky mohou být objektem následujícího kvalitativního uvažování.

5.5 Prostorové hierarchické shlukování

Carvalho a kol. ve své práci [7] studovali metodiku pro hierarchické prostorové shlukování. Zkoumali, jak se liší výsledky shlukování s platným politickým rozdělením Brazílie. Pro shlukování využily algoritmus a techniky, které jsou popsány níže.

5.5.1 Algoritmus prostorového hierarchického shlukování

1) Nechť C je počáteční databáze N územních jednotek. Zpočátku, každé z těchto N pozorování je izolované seskupení a má sadu atributů (proměnných) $[x_{i,1}, x_{i,2}, ..., x_{i,m}]$. Pro každou z těchto N jednotek je nutné určit seznam sousedních objektů, v závislosti na některém prostorovém kritériu. V práci byly zkoumány dvě definice okolí. V prvním

případě byla brazilská města považována za sousedy, pokud měly alespoň jednu hranici společnou - tento druh a okolí je známý v literatuře o prostorové statistice jako „okolí typu věž“. V druhém případě byla považována města za sousední, pokud měly společný alespoň jeden hraniční bod - tento typ okolí je znám jako „okolí typu královna“. Je zřejmé, že „okolí typu královna“ je méně omezující než „okolí typu věž“. Nicméně, jak ukázaly výsledky, rozdíly ve výsledcích jsou zcela zanedbatelné.

2) Spočítá se vektor vzdáleností (ne geografických, ale mezi vektory proměnných/indikátorů) mezi všemi páry tvořenými přísně sousedními prvky ze seznamu jednotek *N* a vybuduje se matice blízkosti (symetrická).

3) Nechť *I* a *J* se dvěma zeměpisnými jednotkami představují nejmenší vzdálenost vektoru mezi nimi, s tím omezením, že *I* a *J* jsou sousedi. Tento pár *I* a *J* se seskupí do jednoho shluku. Počet shluků bude tedy *N-1*.

4) V nejjednodušším případě je pro definici nového shluku, který tvoří celky *I* a *J*, nutné kombinovat seznamy sousedů. Proto bude nový seznam sousedů vytvořen spojením seznamu sousedů města *I* a seznamu sousedů města *J* (sjednocení obou relací).

5) Pro nových *N-1* shluků, po spojení, jak je popsáno v kroku 3, musí být aktualizována matice blízkosti. Aktualizace matice blízkosti (nebo vzdálenosti) závisí na metodě shlukování. Například pro metodu nejbližšího souseda je vzdálenost mezi dvěma shluky *I* a *J* minimální vektor vzdáleností mezi všemi dvojicemi vektorů proměnné ve dvou shlucích. Na druhou stranu pro metodu nejvzdálenějšího souseda je vzdálenost mezi dvěma shluky maximální vektor vzdáleností mezi všemi páry vektorů.

6) Opakujeme kroky 3 až 5, dokud zbude jen jeden shluk, který bude obsahovat všech *N* originálních zeměpisných jednotek.

Na konci procesu je strom, který vzniká v případě tradičních hierarchických shlukovacích metod, který charakterizuje seskupování, které se konalo v každém kroku algoritmu. Opět platí, že výzkumník může využít některý z tradičních ukazatelů (např. *CCC*, *pseudo-F* a *pseudo-t^2*) pro výběr nejvhodnějšího počtu skupin. Nicméně algoritmus, který Carvalho a kolektiv v dokumentu používá, má podstatné změny oproti tradičním hierarchickým shlukovacím algoritmům. Díky tomu vlastnosti těchto statistických ukazatelů se nutně nemusejí shodovat s tradičním shlukováním (neprostorovým), což naznačuje, že je třeba pozdější studium o chování těchto ukazatelů. Na druhou stranu

přímé použití statistických kritérií nemusí nutně vést k takovému počtu shluků, které dává smysl v souladu s cíli každého studia.

5.5.2 Kritéria pro výběr počtu shluků

Podle Carvalha a kolektivu [7] je jeden z nejpoužívanějších indexů Sarleho *CCC* (Cubic Clustering Criterion), který v neprostorových hierarchických algoritmech testuje nulovou hypotézu H_0, že údaje jsou vzorkem z rovnoměrného rozložení oproti alternativní hypotéze H_1, že údaje jsou vzorkem z více sférických vícerozměrných normálních distribucí se shodnými rozptyly a různou pravděpodobností. Kladné hodnoty *CCC* způsobí zamítnutí nulové hypotézy H_0

Vztah pro *CCC* je dán následovně:

$$CCC = \log\left[\frac{1 - E\left[R^2\right]}{1 - R^2}\right] \cdot v \qquad (15)$$

kde v je počet proměnných v databázi, R^2 je index determinace, $E[R^2]$ je očekávaný index determinace a *log* je přirozený logaritmus. Vztah pro index determinace a očekávaný index determinace je prezentovaný v práci Sarleho [33].

Další užívaná kritéria jsou *pseudo-t*^{2}, *semipartial-R*^{2} *(SPRSQ)* a *pseudo-F*. Vysoké hodnoty pro *pseudo-F* naznačují, že průměrný vektor každého shluku je odlišný, to znamená, že každý shluk významně závisí na konfiguraci. Proto jedním ze způsobů použití *pseudo-F* je srovnání vysokých hodnot v grafu *pseudo-F* oproti počtu shluků; zvolený počet shluků odpovídá vrcholu *pseudo-F* grafu. Na druhou stranu, pokud je *pseudo-t*^{2} statistika vysoká na určité úrovni procesu spojování dvou shluků, pak by tyto shluky neměly být spojeny, protože jejich průměrné vektory můžou být považovány za rozdílné.

Proto současná literatura doporučuje, abychom při hledání vrcholových hodnot na sekvenci *pseudo-t*^{2} statistiky použili počet shluků bezprostředně nad počtem shluků, které odpovídají vrcholu.

Semipartial-R^2 kritérium se vypočte jako pokles hodnoty indexu determinace, způsobený spojením dvou shluků *(C_k a C_l)*, tj.

$$SPRSQ(C_k) = R^2(C_1) - R^2(C_k) \qquad (16)$$

Nízké hodnoty naznačují, že změna meziskupinové variability (a tudíž i vnitroskupinové) je menší a tudíž lze dva shluky považovat za jeden. Naproti tomu vysoké hodnoty signalizují, že shluky jsou pravděpodobně odlišné.

Sociálně-ekonomická situace brazilských měst se popisovala na základě těchto indikátorů:

a) Míra nezaměstnanosti v obci (zaměstnaná populace dělená celkovou populací)
b) Procentuální podíl městské populace na území města
c) Demografické proměnné: míra dlouhověkosti a porodnosti obci v roce 2000
d) Městská infrastruktura a stav bydlení: procento sídel s veřejným osvětlením, identifikace (PSČ), kanalizace, vodovod, dlážděné ulice, energetické a odpadové hospodářství
e) Studijní výsledky: procento dětí od 5 do 6 let ve škole; procento dětí od 7 do 14 let s přístupem na základní školu, procento dospívajících od 15 do 17 let s přístupem na střední školu; procento obyvatel od 18 do 24 let s přístupem k vyššímu vzdělání; procento dětí od 7 do 14 let s více než jedním rokem zpoždění ve vzdělávání; procento učitelů na základní škole s vyšším vzděláním; průměr let vzdělávání pro obyvatele od 25 let a starších; procento negramotných lidí od 15 let a starších; procento negramotných lidí mezi 10 a 14 lety
f) Příjem sídla na jednoho obyvatele, podíl příjmů na hlavu odvozený z pracovních příjmů, procento příjmů na hlavu z vládní pomoci v roce 2000
g) Procento chudých lidí v obci v roce 2000
h) Nerovnost příjmů (Giniho index) v roce 2000

i) Proměnné týkající se veřejného zdraví: úmrtnost dětí do 1 roku věku a do 5 let v roce 2000 a pravděpodobnost přežití 60 let v roce 2000.

j) Míra vražd v obci v roce 2002

Proces postupné agregace shluků byl omezen podmínkou, že musí vzniknout minimálně tři shluky (pokud zde nebylo žádné jiné sousedství, ať už se jednalo o „okolí typu věž“ nebo o „okolí typu královna“).

Výsledky ukázaly, že metoda nejbližšího souseda, metoda průměrová, metoda těžiště a metoda mediánová mají tendenci tvořit velice odlišné shluky, pokud jde o počty geografických jednotek. Metoda nejvzdálenějšího souseda a zejména Wardova metoda jsou ty, které mají tendenci vytvářet shluky s nejvíce homogenní velikostí – tato skutečnost byla vzhledem k výpočtu Wardovy metody očekávaná. Pokud jde o počet geografických jednotek, z pěti studovaných vzdáleností nejvíce homogenní shluky poskytuje vzdálenost typu Manhattan.

„Sousedství typu věž“ nebo „sousedství typu královna“ jsou si velmi podobné co do počtu obcí v jednotlivých shlucích, tudíž vliv na volbu typu sousedství má velice zanedbatelný vliv.

Výsledky získané sociálně-ekonomickým hierarchickým shlukováním byly porovnány s výsledky stávajícího politického uspořádání pomocí indexu *TCSS*, který popisuje rozptyl uvnitř shluku (součet čtverců odchylek od střední hodnoty každého shluku).

Výraz pro *TCSS* je dán

$$TCSS = \sum_{k=1}^{G} \sum_{i \in C_k} \sum_{l=1}^{v} \left[x_{k,i,l} - \bar{x}_{k,l} \right]^2 \qquad (17)$$

kde v je celkový počet proměnných v databázi, G je počet shluků (podle politického uspořádání je G = 27, 137 nebo 558), C_k je soubor obcí ve shluku, k, $x_{k,i,l}$ jsou proměnné l v obci, i a $\bar{x}_{k,l}$ je průměr proměnné l v rámci shluku k. Aby bylo možné porovnat různé metody shlukování a politického rozdělení, jsou vykazované hodnoty relativní variabilitou každé metody proti variabilitě pro odpovídající politické rozdělení.

V tomto případě je relativní variabilita $\Delta TCSS_{Metoda}$ je dána:

$$\Delta TCSS_{Metoda} = 100 \; x \; TCSS_{Metoda}/TCSS_{Politické\ rozdělení} \qquad (18)$$

Při zkoumání prezentovaných výsledků pro různé metody shlukování a různé vektory vzdáleností je jasně patrné, že shluky vytvořené použitím shlukovací Wardovy metody mají nejnižší celkovou variabilitu než politické rozdělení. To platí pro mikroregiony (558 shluků), mezoregiony (137 shluků) a státy (27 shluků). Při použití Wardovy metody se variabilita v případě mikroregionů sníží téměř o 50 %. Při použití metody nejvzdálenějšího souseda je obecně oproti politickému rozdělení variabilita také nižší; výjimka je při použití vzdálenosti Mahalanobis. U ostatních metod jsou výsledky méně povzbudivé; v případě států a mezoregionů je variabilita větší než u politického rozdělení. Pro mikroregiony je variabilita menší než u politického rozdělení.

6 FAKTOROVÁ ANALÝZA

Poprvé se objevila v psychologii (Charles Spearman ji použil pro testy inteligence), ale poté našla široké využití v sociologii, marketingu a dalších oblastech. Primární funkcí faktorové analýzy je vysvětlení rozptylu pozorovaných proměnných pomocí menšího počtu latentních proměnných – tzv. faktorů (redukce proměnných). Smyslem je tedy měřit něco, co není měřitelné přímo. [35]

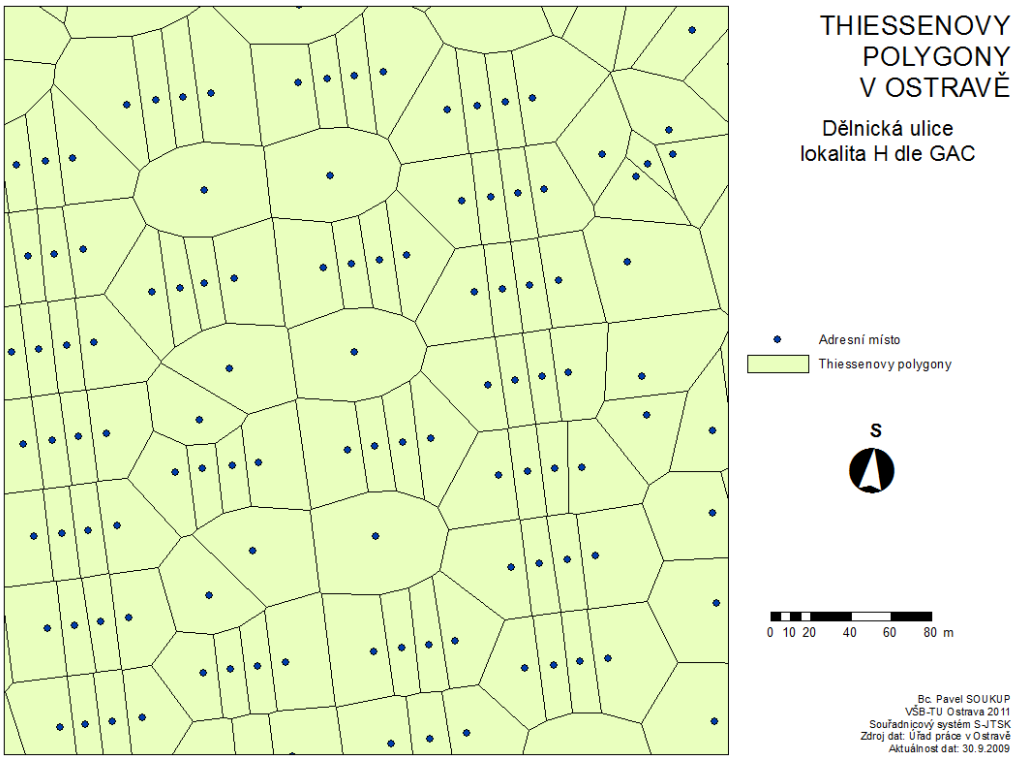
Ve faktorové analýze se předpokládá, že každá vstupující proměnná se může vyjádřit lineární kombinací nevelkého počtu společných skrytých faktorů a jediného chybového faktoru. [26] Počet použitých faktorů může být různý; čím více faktorů, tím je vysvětleno větší procento rozptylu proměnných. Na druhou stranu smyslem faktorové analýzy je nalézt pokud možno co nejmenší přijatelný počet faktorů. Proto je třeba počet hledaných faktorů určovat podle konkrétních dat. Můžeme použít teoretický předpoklad na základě znalosti zkoumané látky nebo například Kaiserovo pravidlo (vlastní číslo faktoru vyšší než 1).

Podmínkou pro faktorovou analýzu je použití kardinálních proměnných nebo ordinálních proměnných se škálou minimálně pěti hodnot. Ze zjištěných faktorů lze vytvořit nové proměnné a dále s nimi statisticky pracovat.

Cílem faktorové analýzy je v tomto případě zjištění skrytých proměnných, faktorů, které působí na trh práce a vazbu jednotlivých proměnných (společně vystupujících v jednom faktoru).

7 THIESSENOVY POLYGONY

Thiessenovy polygony (nebo také Voronovy nebo Dirichletovy polygony nebo proximální zóny) jsou polygony, jejichž hranice definují oblast nejbližší každému bodu vzhledem ke všem dalším bodům a definují jednotlivé oblasti. Tyto budou použity při definování sousedství v hierarchickém prostorovém shlukování. Jako soused bude brán adresní bod, který má se sousedním bodem alespoň jeden společný bod na hraně Thiessenova polygonu. Hranice polygonů jsou stejně vzdáleny k ohniskovým objektům. V rastrovém datovém modelu se pixelu přidělí hodnota nejbližšího známého bodu. [18]



Obr. 3.: Thiessenovy polygony vystavěné nad adresními body v lokalitě Dělnická

8 ZDROJE DAT

Pro tuto diplomovou práci bylo třeba mít k dispozici rozsáhlý a dostatečně důvěryhodný datový soubor, jelikož jsem se snažil dosáhnout toho, aby výsledky byly dostatečně relevantní. Jako takový se nejlépe jevil Registr sčítacích obvodů a budov (dále jen RSO), který vydává Český statistický úřad (ČSÚ), protože poskytuje bodovou lokalizaci adres a obsahuje dobře propracovaný a aktualizovaný systém územně-správního uspořádání a zejména databázi adresních míst. RSO je použito také proto, že zpracovává data ze SLDB, která by každý očekával, jsou již značně zastaralá, jelikož poslední sčítání proběhlo v roce 2001. Další sčítání proběhne ve dnech 26. 3. - 14. 4. 2011.

Další použitá data pocházejí z Magistrátu města Ostravy. Propojením anonymizovaných údajů, které poskytl Úřad práce v Ostravě, s daty RSO vznikly další vrstvy, které byly v práci použity.

8.1 Adresní místa

Agenda RSO obsahuje (kromě jiného) bodovou vrstvu adresních míst. Jejich struktura je popsána v tab. 1. Já jsem pracoval s vyexportovanou částí, která se vztahovala pouze pro město Ostrava. Vrstva obsahuje lokalizaci adresních míst Ostravy ve formě adresního bodu [9] a ve třech datových sadách, které jsou aktuální k 1. lednu 2009, k 1. dubnu 2009 a k 1. červenci 2009 (používaná datová sada má 29956 záznamů).

Souřadnice jsou udávány v S-JTSK. K vytvoření datové sady byly použity různé zdroje - pokud byla jako podklad použita digitální katastrální mapa (DKM) je udávána polohová přesnost 0,5 metru (týká se všech katastrálních území Ostravy). Pokud by šlo o ruční digitalizaci, udávaná polohová přesnost by pak byla 1 metr [8].

V tabulce č. 2 je vysvětlen obsah pole „charakter změny“, který je vyjádřen kódem.

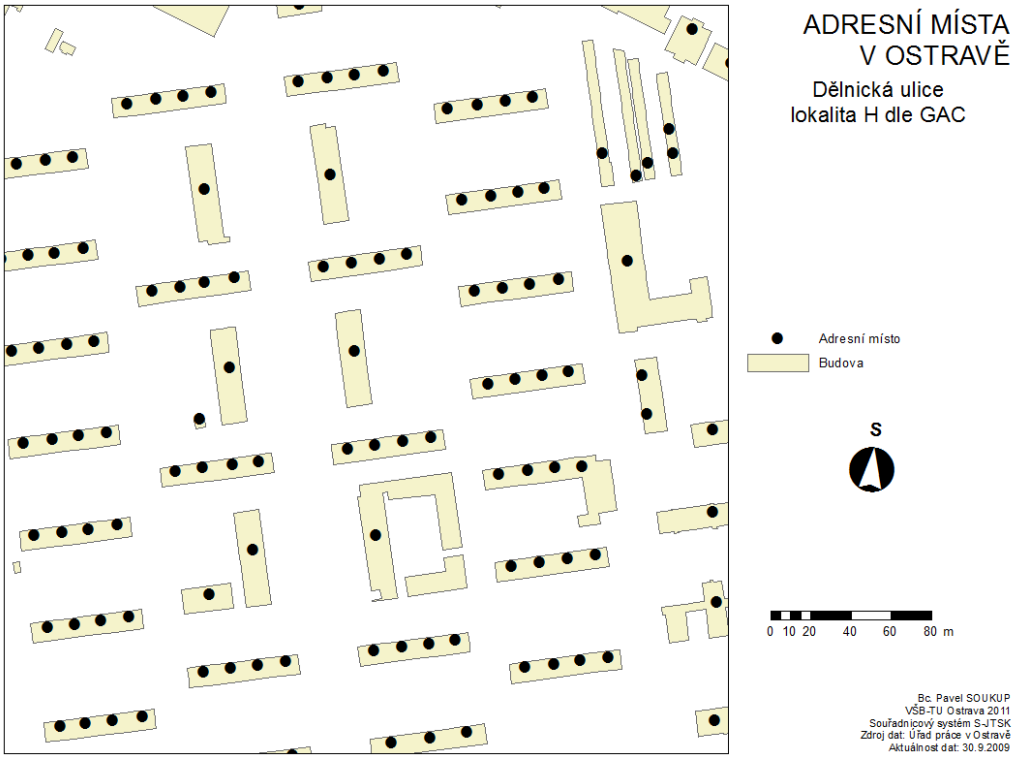
Tab. 1.: Sada atributů bodové vrstvy Adresních míst RSO ČSÚ – upravené podle [8]

Typický název	Slovní popis (sémantika)	Datový typ
IDADR	unikátní identifikátor adresy v ČR	Text (11)
ADRESA_KOD	kód adresy dle ÚIR-ADR (pouze u vícenásobných adres)	Double
ZMENA	charakter změny v datové sadě	Short Integer
DAT_ZPRAC	datum zpracování datové sady, ke kterému jsou změny vztaženy	Date
ADR_JTSK_X	souřadnice adresního místa X	Double
ADR_JTSK_Y	souřadnice adresního místa Y	Double
VICEADR	příznak násobné adresy	Short Integer
ZDROJ	zdroj lokalizační informace	Text (10)
IDOB	unikátní identifikátor budovy v ČR	Text (10)
PC_BUDOV	pořadové číslo budovy	Short Integer
TYP_CIS	typ domovního čísla (popisné, evidenční, náhradní)	Short Integer
CIS_D	domovní číslo	Long Integer
CIS_O	číslo orientační v rámci ulice a veřejného prostranství	Text (4)
ULICE_ID	jedinečný identifikátor ulice v České republice	Text (7)
NAZEV_UL_A	název ulice adresní (velká i malá písmena)	Text (40)
PSC	poštovní směrovací číslo dodávací pošty	Text (5)
NAZ_POSTA	název dodávací pošty	Text (30)
NAZ_NNUTS4	název okresu (NUTS4), resp. LAU1	Text (40)
KOD_OBEC	kód obce	Text (6)
NAZ_OBEC	název obce	Text (40)
KOD_CAST_D	kód části obce	Text (6)
NAZ_CAST_D	název části obce	Text (40)
KOD_KU_A	kód katastrálního území	Text (6)
NAZ_KU_A	název katastrálního území	Text (40)
IDSO	jedinečný identifikátor sčítacího obvodu v ČR	Text (6)
TYP_ADRESA	rozlišení adresy dle její váhy	Short Integer
ZAD_VCHOD	příznak zadního vchodu (ANO = 1)	Short Integer
CUZKBUD_ID	umělý identifikátor budovy v ISKN	Double
KOD_UZOHMP	bezvýznamový kód územního obvodu hlavního města Prahy	Text (3)
NAZ_UZOHMP	název územního obvodu hlavního města Prahy	Text (10)
LAU1	klasifikace LAU, textová hodnota kódu okresu	Text (6)
PCD	unikátní identifikátor adresního místa v systému ISEO-Adresa	Double
PARCELA	kód parcelního čísla	Text (10)

Poznámka: použité datové typy podle Microsoft Access verze 2000, resp., MS Jet Engine 4.0.

Tab. 2.: Číselník změn pro vrstvu adresních míst

Kód změny	Charakter změny
11	nová budova - nový definiční bod
21	oprava/zpřesnění lokalizace budovy - výměna definičního bodu
22	změna přirozené identifikace (PI) budovy (kombinace čísla domovního, typu č.d. a části obce) - změna PI
23	oprava/zpřesnění lokalizace spolu se změnou PI - výměna definičního bodu
31	zrušení budovy (pouze u aktualizačních balíčků) - výmaz definičního bodu
41	doplnění lokalizace k budově - nový definiční bod
51	chybná lokalizace - probíhá náprava (pouze u aktualizačních balíčků) - výmaz definičního bodu
61	budova má v popisné části registru statut rozpracovaných - není součástí exportů popisných dat a má-li budova tento statut v době vydání datové sady, není její součástí; tyto budovy jsou průběžně ověřovány a v závislosti na výsledku šetření jsou buď vyloučeny z evidence nebo znovu plnohodnotně zaevidovány (v tomto případě se vracejí do další verze datové sady s kódem 69)
69	oživení rozpracované budovy (viz 61)
91	beze změny (u stavových datových sad)/ v průběhu zpracovatelského období došlo k manipulaci (u aktualizačních balíčků)



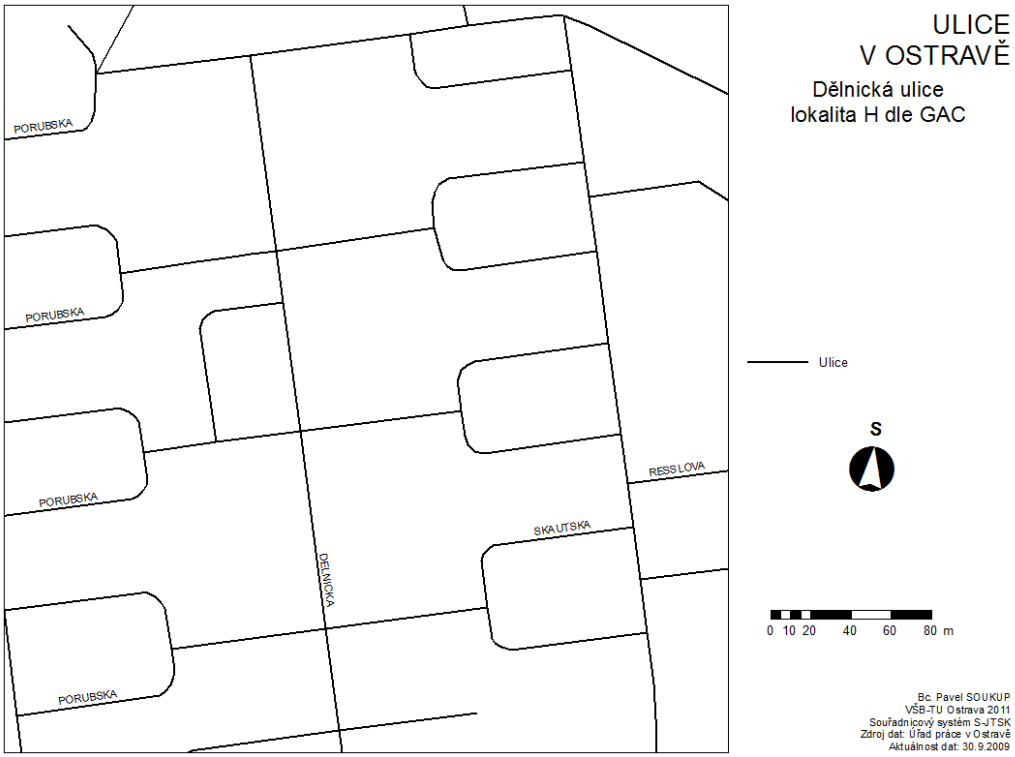
Obr. 4.: Adresní místa v Ostravě v lokalitě Dělnická

8.2 Uliční síť

Liniovou vrstvu uliční sítě poskytl Magistrát města Ostravy (MMO) a obsahuje 16135 záznamů. Data jsou aktuální k 1. lednu 2009. Souřadnice jsou udávány v S-JTSK.

Tab. 3.: Sada atributů liniové vrstvy ULICE

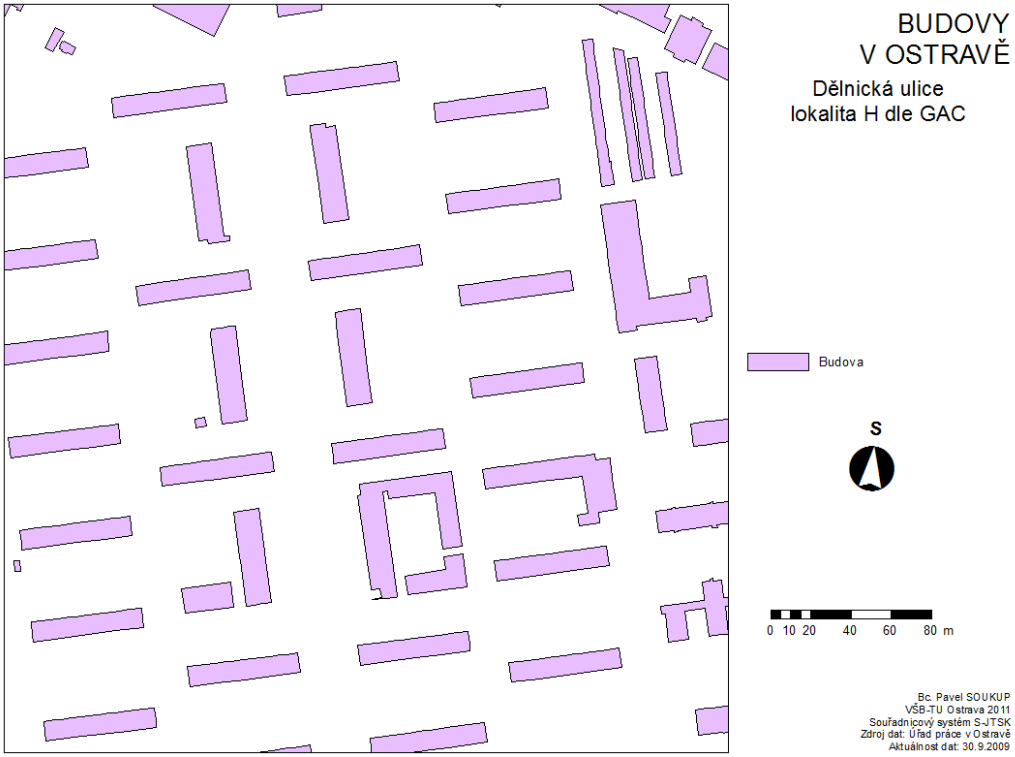
Typický název	Slovní popis (sémantika)	Datový typ
LENGTH	Délka úseku linie v metrech	Double
ULICE_	Identifikátor ulice ve vrstvě	Double
FROM_L	Začátek číslování na levé straně linie	Short Integer
TO_L	Konec číslování na levé straně linie	Short Integer
FROM_R	Začátek číslování na pravé straně linie	Short Integer
TO_R	Konec číslování na pravé straně linie	Short Integer
STREET_NAM	Název ulice s diakritikou	Text (40)
STR_NAME	Název ulice bez diakritiky	Text (40)
CIS_ULI	Číslo ulice	Long Integer
KOD	Rozlišení ulice dle provozu	Short Integer
JMENO_UL	Název ulice původní	Text (40)
NAME_ASCI7	Název ulice pomocí znaků ASCII	Text (40)
ONEWAY	rozlišení ulice dle provozu	Text (5)
TRIDA	třída komunikace	Text (2)
CISLO_KOM	číslo komunikace	Text (12)
SHAPE_LEN	délka linie v metrech	Double
SHAPE_FID	jedinečný identifikátor v rámci databáze	Long Integer



Obr. 5.: Uliční síť v Ostravě v lokalitě Dělnická

8.3 Budovy

Polygonovou vrstvu budov poskytl MMO a nachází se v ní 53799 záznamů. Vrstva obsahuje všechny stavby, které jsou pevně spojeny se zemí. Aktuální jsou k 1. lednu 2009. Souřadnice jsou udávány v S-JTSK.



Obr. 6.: Budovy v Ostravě v lokalitě Dělnická

Tab. 4.: Sada atributů polygonové vrstvy BUDOVY

Typický název	Slovní popis (sémantika)	Datový typ
POPIS	popis objektu	Text (10)
SHAPE_AREA	plocha objektu	Double
SHAPE_LEN	délka hraniční linie objektu	Double
SHAPE_FID	jedinečný identifikátor objektu	Long Integer

8.4 Adresní body z úřadu práce

Na základě spolupráce s Úřadem práce v Ostravě (dále jen ÚP Ostrava) byl proveden export vybraných anonymizovaných údajů o registrovaných uchazečích o zaměstnání vždy k danému datu. Na základě propojení s vrstvou adresních bodů z RSO a adresními body MMO bylo provedeno geokódování adres a připravena bodová vrstva s označením EvidenceUP. Vrstva obsahuje 18745 záznamů a je aktuální k 30.9.2009.

Tab. 5.: Sada atributů bodové vrstvy EVIDENCEUP

Typický název	Slovní popis (sémantika)	Datový typ
OBC_IDCISL	Umělý identifikátor	Text (16)
X	Souřadnice X	Double
Y	Souřadnice Y	Double
POCET	Počet záznamů	Double
DEVX	Směrodatná odchylka v souřadnici X	Double
DEVY	Směrodatná odchylka v souřadnici Y	Double

8.5 Agregované demografické a ekonomické údaje pro adresní body

Datová vrstva AgregaceUP byla zpracována v rámci spolupráce s Úřadem práce v Ostravě. Jedná se o bodovou vrstvu, která obsahuje již zpracované, agregované informace o věkovém složení obyvatel a situaci v nezaměstnanosti v Ostravě. Je v ní 7354 záznamů. V tabulce číslo 6 je uvedena struktura dat, která jsem převzal pro další využití. Data jsou aktuální k 30. září 2009. Souřadnice jsou udávány v S-JTSK.

Tab. 6.: Sada atributů bodové vrstvy AGREGACEUP

Systémový název	Slovní popis	Datový typ
ID	Jedinečný identifikátor bodu	Text (5)
X	Souřadnice X bodu	Double
Y	Souřadnice Y bodu	Double
POCET	Počet adres přiřazených k dotyčnému bodu	Double
DEVX	Směrodatná odchylka v souřadnici X	Double
DEVY	Směrodatná odchylka v souřadnici Y	Double
UC	Počet uchazečů o zaměstnání	Double
UC0024	Počet uchazečů o zaměstnání do 25 let	Double
UC5099	Počet uchazečů o zaměstnání nad 50	Double
UCVABC	Počet uchazečů o zaměstnání s nízkým vzděláním (A – bez vzdělání, B – nedokončené základní vzdělání, C – dokončené základní vzdělání).	Double
UCZPS	Počet uchazečů o zaměstnání se změněnou pracovní schopností	Double
UCE12	Počet uchazečů o zaměstnání v evidenci déle než 12 měsíců	Double
O	Počet osob	Double
OPV	Počet osob v produktivním věku (15 – 65 let)	Double
O0024	Počet osob do 25 let	Double
O5099	Počet osob nad 50 let	Double
PUC_OPV	Podíl počtu uchazečů na osoby v produktivním věku [%]	Double
PC0025_O	Podíl osob do 25 let [%]	Double
PC5099_O	Podíl osob ve věku nad 50 let [%]	Double
PCVABC_U	Podíl uchazečů s nízkým vzděláním z celkového počtu uchazečů [%]	Double
PCZPS_U	Podíl uchazečů se změněnou pracovní schopností z celkového počtu uchazečů	Double
PCE12_U	Podíl uchazečů v evidenci nad 12 měsíců z celkového počtu uchazečů [%]	Double
TERMIN	Datum aktualizace	Date
UC1824	Počet uchazečů o zaměstnání mezi 18 – 25 let	Double
UC5064	Počet uchazečů o zaměstnání mezi 50 – 65 let	Double
O1824	Počet osob ve věku 18 – 25 let	Double
O5064	Počet osob ve věku 50 – 65 let	Double

Prohlášení o utajeném obsahu závěrečné práce

Diplomová práce s názvem

Prostorové shlukování v městském prostředí

autora Bc. Pavla Soukupa

obsahuje citlivé informace k vymezování a popisu sociálně problémových lokalit v Ostravě. Tato závěrečná práce obsahuje informace, které jsou předmětem ochrany poskytovatele nebo ochrany vyplývající z platných zákonů a proto ji nelze obecně zpřístupnit v elektronickém systému prostřednictvím databáze kvalifikačních prací ani ji nelze volně zpřístupnit ve veřejně přístupných knihovních fondech. Celá diplomová práce je uložena na Institutu geoinformatiky.

V Ostravě dne 24. 5. 2011

.....

Doc. Dr. Ing. Jiří Horák

Vedoucí DP

11 DISKUZE VÝSLEDKŮ

Indexy prostorové heterogenity nám dávají orientační náhled na rozložení sledovaných proměnných již na úrovni základního administrativního členění. Pokud máme relevantní data (dvě nepřekrývající se skupiny), je jejich výpočet snadný a dá nám zevrubný náhled na situaci. Jejich nevýhoda je ovšem v tom, že data musí příslušet k nějaké městské části a tento údaj v primárních datech chybí. Naštěstí to díky překryvným operacím není tak složité. Dalším velkým problémem je to, že tyto indexy hodnotí Ostravu jako celek – jestli je nebo není segregována, apod. (zajímavé by spíše bylo srovnání s jinými městy). Bohužel ke zjištění, ve kterých lokalitách se nezaměstnaní segregují nejvíce, se indexy prostorové heterogenity nehodí.

K tomu, abychom zjistili prostorový trend nezaměstnaných, je vhodné zkoumat data v kontinuální podobě. Pro studium dat v kontinuální podobě (aby lépe vynikly rozdíly), byla provedena interpolace. Na mapách, které vznikly interpolací, jsou již jasně patrné lokality s vysokými hodnotami sledovaných proměnných. Ovšem je nutné dát si pozor na rozsáhlé oblasti, kde není osídlení a my nevíme, jakou hodnotu zde nezaměstnanost má, zkrátka zde není definován ukazatel. V těchto místech tedy hodnoty neznáme, jedná se pouze o určitý spekulativní odhad, který je výrazně méně spolehlivý. Mezi hlavní nevýhody interpolace patří problém volby interpolační techniky, která má značný vliv na výsledek (větší či menší vyhlazení, problém určení hranic, atd.). Z výsledků interpolace ovšem nezjistíme strukturu hodnot v jednotlivých lokalitách.

Abychom mohli zkoumat rozložení hodnot přímo detailně v jednotlivých vytipovaných lokalitách, byla zvolena metoda bodového kartodiagramu. Jeho výhodou je jednoznačná interpretace hodnot sledovaných proměnných.

Vizualizace výsledků analýzy LISA nám též dává detailní náhled na jednotlivé lokality. Můžeme již vidět zřetelné shlukování. Analýza LISA umí dobře odhalit shluky vysokých nebo nízkých hodnot. Nevýhodou analýzy LISA ovšem je to, že sice odhalí shluky vysokých nebo nízkých hodnot, ale to může být dáno i tím, že tyto vysoké (nízké) hodnoty jsou obklopeny nízkými (vysokými) hodnotami. Analýza LISA tedy odhalí shluky anomálních hodnot, ale její výsledky jsou značně ovlivněny okolím.

Další metoda, metoda hierarchického prostorového shlukování již okolím ovlivňována není. Tato metoda spojuje do shluků adresní body na základě jejich podobnosti (při různé definici sousedství). Při porovnání výsledků hierarchického prostorového shlukování s mnou vytvořenými shluky (které jsem si předtím vytvořil ručně), se jako lepší varianta jeví použití euklidovské vzdálenosti. Sousedství, které je definováno pomocí Thiessenových polygonů totiž nespojí do shluku adresní místa, která viditelně patří do jednoho shluku, ale díky jejich topologii nejsou brány jako sousední body.

Pro měření heterogenity a shlukování v území bych doporučil udělat analýzu LISA, která dobře odhalí shluky vysokých hodnot nebo metodu hierarchického prostorového shlukování se sousedstvím definovaným euklidovskou vzdáleností s mezní hodnotou 50 metrů. Ještě by bylo dobré vyzkoušet Wardovu metodu namísto používané metody nejbližšího souseda (dle odborné literatury poskytuje nejlepší výsledky).

12 ZÁVĚR

Z výpočtů heterogenity území města Ostravy bylo orientačně zjištěno, že k mírnější segregaci či shlukování uchazečů o zaměstnání dochází již na úrovni základního administrativního členění Ostravy (městské části). Nejvyšších hodnot dosahoval index koncentrace, což nám napovídá o tom, že nezaměstnaní žijí v oblastech hustší zástavby (více lidí na menší ploše). Indexy odlišnosti a segregace dosahovaly nejvyšších hodnot pro sledované proměnné počtu dlouhodobě nezaměstnaných a počtu nezaměstnaných s nízkým vzděláním, z čehož můžeme vyvodit, jak a kde které skupiny obyvatel (potažmo nezaměstnaných) bydlí (jak jsou skupiny separovány).

Při mapování přesného průběhu hodnot ve vytipovaných lokalitách můžeme mluvit též o shlukování adres s podobnými hodnotami sledovaných proměnných. Ukázalo se, že adresy, které jsou k sobě blíže, mají podobné hodnoty sledovaných proměnných než adresy, které jsou dále od sebe. Na mapách, které vznikly interpolací, jsou jasně vymezené lokality s vyšší koncentrací sledovaných proměnných.

Z porovnání výsledků prostorové autokorelace bylo zjištěno, že nejvýznamnější prostorová asociace se nachází u proměnné počtu nezaměstnaných a u proměnné počtu uchazečů o zaměstnání s nízkým vzděláním. Je zajímavé sledovat, jak se hodnota Moranova I kritéria s různými parametry sousedství mění – nejvyšší hodnotu má při mezní vzdálenosti 50 metrů (když se v potaz bere čistá euklidovská vzdálenost, nikoliv k-nejbližších sousedů. To je dáno tím, jaké budovy se nacházejí tak blízko u sebe (jsou to rodinné domky a v nich je mnohem menší variabilita v kapacitě i počtu lidí).

Analýza LISA identifikovala i další lokality, o kterých v mnou studované odborné literatuře není žádná zmínka. Analýza LISA odhalila shluky nadprůměrných hodnot a podprůměrných hodnot. Analýzou LISA tak byly identifikovány v podstatě jádra či centra koncentrace nezaměstnaných. Výsledky byly opět porovnány s vytipovanými lokalitami a zkoumala se shoda. Pro sledované proměnné nastal největší soulad u proměnné počtu uchazečů s nízkým vzděláním a u proměnné počtu uchazečů, kteří jsou v evidenci déle než 12 měsíců.

Vlastní hierarchické prostorové shlukování implementovalo pouze euklidovský typ vzdálenosti a sousedství „typu královna“, které bylo vymezeno na základě Thiessenových polygonů a jako metodu shlukování používá metodu nejbližšího souseda. Výsledky ukazují, že lepší shluky jsou generovány při použití euklidovského typu vzdáleností s limitním parametrem 50 metrů (do jaké míry je adresní místo bráno jako soused). Výsledky vypadají logičtěji, protože Thiessenovy polygony definují sousedství i místům, které spolu v reálném světě nesousedí (jsou třeba odděleny dalším domem).

13 RESUME

This thesis deals with the spatial clustering in the urban areas. My main goal is to compare selected methods that can be used for the studies of spatial clustering (spatial heterogenity indexes, spatial autocorrelation measurings with the help of Moran's I criterion, the LISA analysis and the spatial hierarchical clustering itself). The data come from the Employment Office in Ostrava (I have preprocessed data localized right to the address points).

From the measure of heterogenity in the area of the city of Ostrava it was tentatively found out that a mild segregation or clustering of the job seekers is extensible, identifiable already for the aggrefation level of the Ostrava town districts. To be able to compare whether the results changes in time, I calculated all the indexes in two time period. The results showed that the spatial heterogenity indexes data almost did not change during the observed year. The heterogenity indexes was calculated for following data: number of unemployed, number of unemployed with a low degree of education and the number of long-term unemployed more than 12 months. The concentration index reached the highest values of all the observed variables, from which we can see that the unemployed mostly live in densely built-up areas (more people on a smaller area). The dissimilarity and interaction indexes reached similar values for all three variables (about 0,2), from which we can see how and where different population groups (or unemployed groups) live. The isolation index cannot be interpreted from any of the observed variables (the value reached 0,1 – at the number of unemployed with a low degree of education). From the comparison of the result I would accentuate the segration index value for the number of unemployed with a low degree of education that reached a truly high value (0,52 and 0,53 in 2009 and 2010 respectively) – this index shows how the groups are mutually separated. A value higher than 0,5 tells us how many people with a low degree of education would have to move within the whole city to reach an even distribution of the population. In this case it would be more than 50 % of all the unemployed with a low degree of education (in the absolute number it would be roughly more than 3,500 unemployed with a low degree of education from a total number of 7,000). The big

problem is that these indexes evaluate Ostrava as whole. Spatial heterogenity indexes does not fit to determinate in which locations the unemployed segregate the most.

Then, the factor analysis was made. Its goal was to find hidden variables, factors that influence the labour market and the relations between the particular variables. The meaning of the factor analysis is to compress a large collection of indicators into a tabular number of groups of variables (factors) without losing any important piece of information. Considering my need of the relation analysis and uncovering of the factors used to find a fitting methods for studying heterogenity in the area, I used data aggregated to address points. The factor analysis included 6 factors that explained more than 80 % of the aggregate variability (the population factor, the unemployment factor, the factor of unemployment of young people, the factor of the disadvantaged unemployed, the factor of the unemployed with a low degree of education and the factor of population number).

Before I started to find out the develompent of the values in the area and calculate the spatial autocorrelation, I had to pinpoint certain localities that I could compare the results with. For this purpose I used, *Analýza sociálně vyloučených romských lokalit a absorpční kapacity subjektů působících v této oblasti* (*Analysis of socially excluded Romani localities and absorption capacities of the subjects that operate in this area*) by the GAC Ltd. company. It is an output of a project of the same name that was launched by the Czech Labour Ministry and the Government Council for the Romani Affairs and financed from the European Social Fund and the government budget of the Czech Republic. Within this publication was created a map of socially excluded or with social exclusion endangered Romani localities. Another work that helped me to pinpoit the localities is *Popis sociálně vyloučených romských lokalit v regionu Ostravska* (*Description of the socially excluded Romani localities in the Ostrava Region*) that was ordered by Social Integration Agency and created by Radim Kvasnička. On the basis of these works in picked up 13 localities and compared the results with them.

It is good to observe the value changes in the area continually because the differences in the values will come to the fore. In this work, the development of unemployment (or some of its aspects) in the area of Ostrava is documented, with poits as data bearers. It is necessary to do a space interpolation to get a surface development (so a continual surface). On the maps that was created by interpolation, the localities with a

higher concentration of the observed variables are clearly defined. In the pinpointed areas we can see a clustering of addresses with similar values of the variables. It turned out that the addresses that are closer to each other have more similar values of the observed variables than the addresses that are farther from each other. The variable that shows the most noticeable clustering is the number of job seekers with a low degree of education. From the results of interpolation we not find the structure of values in individual localities.

To examine the distribution of values directly in the detail of selected sites was chosen method of pointed cartographs. For the spatial correlation analysis I used Moran's I criterion and the LISA analysis. Seven social-economical indicators were analyzed. From the results of the spatial autocorrelation I have found out that the most significant spatial association is in the number of unemployed and in the number of job seekers with a low degree of education. It is interesting to watch the changes of the value of Moran's I criterion with different parameters of neighbourhood – its highest value is within the threshold distance of 50 metres (considering a pure Euclidean distance, not the k-closest neighbours). This is because in the distance of 50 metres there are mainly small houses and there is a much lesser variability in it, both in the capacity and in the number of people. Taking the closest neighbourhood into account (the k-nearest neighbours), the most significant spatial correlation appears with using the closest neighbour (with a higher number of neighbours the autocorrelation is lower).

Following the global spacial autocorrelation analysis with the help of Moran's I was made the LISA analysis. The results of the global statistics of the spatial autocorrelation is the assessment of the spatial clustering degree in the whole area. The local statistics (the LISA analysis) allow to identify the areas with a different spatial aurtocorrelation character. For the comparison of the result I chose a weighing scheme with the threshold distance of 50 metres since it looks as the best possibility according to the calculation of Moran's I criterion (the value of Moran's I criterion is highest at this threshold distance). The LISA analysis covered up clusters of both higher and lower than average values. It basically revealed the cores or centres of the concentration of the unemployed. Again, the results were compared with the pinpointed localities and the agreement was checked. The highest harmony for the observed variables appeared in the job seekers with a low degree of education and the number of long-term unemployed more than 12 months. The LISA analysis even identified other localities that were not mentioned in the literature I used.

The LISA analysis reveals clusters of anomalous values, but results are heavily influenced by their neighborhood.

Another method, hierarchical spatial clustering has not affected by neighborhood. The hierarchical spatial clustering itself was implemented as a programme procedure in VBA (Visual Basic for Application) in the database. The procedure itself encompasses only the Euclidean type of distance and the neighbourhood that was defined on the basis of Thiessen polygons. Using the Euclidean type of distance, the point that was within the distance of 50 metres was considered the neighbouring address point. Using the Thiessen polygons, the point that had at least one edge of the Thiessen polygon in common (sometimes this type of neighbourhood is referred as the „queen type“ neighbourhood) were considered the the neighbouring address points. The procedure uses the method of the first neighbour as the method of clustering. This application was tested in three localities. First of all, I defined the clusters by hand and than I compared them with the results made by the programme procedure. The results show that better clusters are generated by using the Euclidean type of distance. The results seem to be more logical because the Thiessen polygon define a neighbourhood even to place that in real world do not adjoin (they are e. g. separated with another house or are truly far from each other).

To measure heterogenity and clustering in the urban areas I would recommend LISA analysis, which well reveals cluster of high values or method of hierarchical clustering with neighbourhood, which is defined by Euclidean distance to the limit value of 50 meters. Another possibility it would be good to try the Ward’s method rather than the used nearest neighbor method (according to scientific literature provide the best results).

SEZNAM POUŽITÉ LITERATURY

1. ANSELIN, Luc. *Community and Environmental Sociology* [online]. 1995. Local Indicators of Spatial Association-LISA The. Dostupné z WWW: <http://www.dces.wisc.edu/documents/articles/curtis/cesoc977/Anselin1995.pdf>.

2. ANSELIN, Luc. *Exploring Spatial Data with GeoDaTM : A Workbook* [online]. Revised Version. University of Illinois : Urbana-Champaign Urbana, 2005. Dostupné z WWW: <http://www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf>.

3. ANSELIN, Luc. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer

Academic Publishers.

4. *Answers.com* [online]. Index of segregation. Dostupné z WWW: <http://www.answers.com/topic/index-of-segregation>.

5. ANTONÍN, Vojáček. *Katedra informatiky a výpočetní technikty ZČU* [online]. 2006. Samoučící se neuronová síť - SOM, Kohonenovy mapy. Dostupné z WWW: <http://www.kiv.zcu.cz/studies/predmety/uir/NS/Samouc_NN2.pdf>.

6. Bailey, T., Gatrell, A.: *Interactive spatial data analysis*. Essex, Longman Scientific & Technical, 1995, 413 s.

7. CARVALHO, Alexandre Xavier Ywata; ALBUQUERQUE, Pedro Henrique Melo; ALMEIDA JUNIOR, Gilberto Rezende de. GUIMARÃES, Rafael Dantas. Clusterização espacial hierárquica. Rev. Bras. Biom [online]. São Paulo : v.27, n.3, p.412-443, 2009. Dostupné z WWW: <http://www.fcav.unesp.br/RME/fasciculos/v27/v27_n3/A6_Alexandre.pdf>.

8. *Český statistický úřad* [online]. Adresní místa. Dostupné z WWW: <http://www.czso.cz/csu/rso.nsf/i/adresni_mista>.

9. *Český statistický úřad* [online]. Co je RSO?. Dostupné z WWW: <http://www.czso.cz/csu/rso.nsf/i/co_je_rso>.

10. *Donald bren school of information and computer science* [online]. Minimum spanning trees. Dostupné z WWW: <http://www.ics.uci.edu/~eppstein/161/960206.html>.

11. GAC (2006) *Analýza sociálně vyloučených romských lokalit a absorpční kapacity subjektů působících v této oblasti*. Projekt Ministerstva práce a sociálních věcí. Dostupný z WWW: <http://www.esfcr.cz/mapa/index.html>.

12. GUILLAIN Rachel., GALLO, Julie, Le. *Measuring Agglomeration: An Exploratory Spatial Analysis approach applied to the Case of Paris and its Surroundings*. 2006.

13. HÁJKOVÁ, Martina. *Prostorové aspekty sociální exkluze - případ Romů*. Olomouc, 2009. 76 s. Diplomová práce. Univerzita Palackého v Olomouci.

14. HARRIS, Richard, LONGLEY, Paul. *Targeting Clusters of Deprivation within Cities.* In STILLWELL, John, CLARKE, Graham. Applied GIS and Spatial Analysis. [s.l.] : John Wiley & Sons, Ltd, 2006. Social Deprivation. s. 88-110. Dostupný z WWW: <http://www3.interscience.wiley.com/cgi-bin/booktext/112468855/BOOKPDFSTART>. ISBN 9780470871331.

15. HOMOLA, Vladimír. *Vysoká škola báňská - Technická univerzita Ostrava* [online]. 10/2002. Interpolace a extrapolace v rovině. Dostupné z WWW: <http://homel.vsb.cz/~hom50/SLBSTATS/IER/GS03.HTM>.

16. HORÁK, Jiří., IVAN, Igor., INSPEKTOR, Tomáš., TVRDÝ Lubor. *Identification and Monitoring of Socially Excluded Localities of Ostrava City using a Register of Unemployment*.

17 HORÁK, Jiří, *Prostorové analýzy dat*. Ostrava: VŠB-TUO, 2008. 158 s.

18. HORÁK, Jiří. *Zpracování dat v GIS*. Ostrava: VŠB-TUO, 2009. 199 s.

19. JARGOWSKY, Paul, A. *Concentration of Poverty and Metropolitan Development*. University of Texas at Dallas. 2006.

20. JARGOWSKY, Paul, A. *Sprawl, Concentration of Poverty, and Urban Inequality*, 2001-06-30, Univerzity of Texas at Dallas.

21. KAŇOK, Jaromír. *Tematická kartografie*. 1. vyd. Ostrava : Ostravská univerzita, 1999. 318 s. ISBN: 80-7042-781-7.

22. KVASNIČKA, Radim. *Popis sociálně vyloučených romských lokalit v regionu Ostravska : Zadavatel: Agentura pro sociální začleňování* [online]. Ostrava : [s.n.], 2010. Dostupné z WWW: <http://www.socialni-zaclenovani.cz/dokumenty/dokumenty-lokality/doc_download/50-popis-socialn-vylouenych-romskych-lokalit-v-regionu-ostravska-kvasnika-r-2010>.

23: MARCUSE, Peter. What's So New About Divided Cities?. *International Journal of Urban and Regional Research*. 17.3.1993, 17, 3, s. 355-365.

24. MARTORI, Joan, Carles., HOBERG, Karen., SURINACH Jordi. *Segregation measures and spatial autocorrelation. Location patterns of immigrant minorities in the Barcelona Region*. 45th Congress of the European Regional Science Association, Vrije Universiteit Amsterdam. 2005-08-23/27.

25. MELOUN, Milan; MILITKÝ, Jiří; HILL, Martin. *Počítačová analýza vícerozměrných dat v příkladech*. Vydání 1. Praha : Nakladatelství Akademie věd České republiky, 2005. 449 s. ISBN 80-200-1335-0.

26. MELOUN, Milan. *Univerzita Pardubice.* Faktorová analýza . Dostupné z WWW: <http://meloun.upce.cz/docs/research/chemometrics/methodology/4dmetody.pdf>.

27. NEZDAŘILOVÁ, Eva. 1984. *Metody kvantitativní analýzy v geografii – se zaměřením na metody regrese a korelace*. Diplomová práce. Praha: Katedra sociální geografie a regionálního rozvoje PřF UK.

28. OSTENDORF, Wim. MUSTERD, Sako., VOS de Sjoerd. *Social Mix and the Neighbourhood Effect. Policy Ambitions and Empirical Evidence*, Housing Studies, Vol. 16, str 371-380, University of Amsterdam 2001.

29. POTTER, Tim. *Effect of Concentration of Poverty at School on Reading Scores*. Planning Research & Evaluation. 2003.

30. REES, Phil, FOTHERINGHAM A. Steward, CHAMPION, Tony. *Modelling Migration for Policy Analysis*. In STILLWELL, John, CLARKE, Graham. Applied GIS and Spatial Analysis. [s.l.] : John Wiley & Sons, Ltd, 2006. NATIONAL SPATIAL PLANNING. s. 258-296. Dostupný z WWW: <http://www3.interscience.wiley.com/cgi-bin/booktext/112468863/BOOKPDFSTART>. ISBN 9780470871331.

31. ROGERSON, Peter. *The Application of New Spatial Statistical Methods to the Detection of Geographical Patterns of Crime*. In STILLWELL, John, CLARKE, Graham. Applied GIS and Spatial Analysis. [s.l.] : John Wiley & Sons, Ltd, 2006. Social Deprivation. s. 151-168. Dostupný z WWW: < http://www3.interscience.wiley.com/cgi-bin/booktext/112468858/BOOKPDFSTART>. ISBN 9780470871331.

32. ŘEZANKOVÁ, Hana. *Vysoká škola ekonomická v Praze* [online]. 2003 [cit. 2011-04-16]. Klasifikace pomocí shlukové analýzy. Dostupné z WWW: <http://nb.vse.cz/~rezanka/shlukova_analyza2003.pdf>.

33. SARLE, W.S. *Cubic clustering criterion*, SAS Technical Report A-108, Cary, NC: SAS Institute, 1983. [online]. Dostupný z WWW: <http://support.sas.com/documentation/onlinedoc/v82/techreport_a108.pdf>.

34. SPURNÁ, Pavlína. *Prostorová autokorelace - všudypřítomný jev při analýze prostorových dat?*. Sociologický časopis. 2008, 44, 4, s. 767-787.

35. ŠKALOUDOVÁ, Alena. *Pedagogická fakulta Univerzity Karlovy v Praze* [online]. 2010 [cit. 2011-04-16]. Faktorová analýza. Dostupné z WWW: <http://userweb.pedf.cuni.cz/kpsp/skalouda/fa/>.

36. TEMELOVÁ, Jana., SÝKORA, Luděk. *Segregace: definice, příčiny, důsledky, řešení*. 20 s. Dostupné z WWW: <http://everest.natur.cuni.cz/akce/segregace/publikace/Temelova_Sykora.pdf>

37. TOUŠEK, Ladislav. *Sociální vyloučení a prostorová segregace.* [online]. dostupné z WWW: < http://www.caat.cz/publikace/44-prehledove-studie/143-socialni-vyloučeni-a-prostorovasegregace>.

38. WISSEN, van Leo. Modelling Regional Economic Growth by Means of Carrying Capacity. In STILLWELL, John, CLARKE, Graham. Applied GIS and Spatial Analysis. [s.l.] : John Wiley & Sons, Ltd, 2006. NATIONAL SPATIAL PLANNING. s. 297-313.

Dostupný z WWW: <http://www3.interscience.wiley.com/cgi-bin/booktext/112468864/BOOKPDFSTART.